GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 13, 2004, 12:02:13 ; Search time 88 Seconds
(without alignments)
12247.843 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 3110
Sequence: 1 agggacgctgccgcaccgcc.....atttccttcaggttttaaaa 1708

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09270437/runat_13072004_121921_9604/app_query.fasta_1.1863
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437_@CGN_1_1_146_@runat_13072004_121921_9604 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2221	71.4	577	11	Q8CGX0	Q8cgx0 rattus norv
2	2216	71.3	577	4	Q9NZI8	Q9nzi8 homo sapien
3	2208	71.0	577	11	O88477	O88477 mus musculu
4	2201	70.8	577	11	Q8BRH1	Q8brh1 mus musculu
5	2201	70.8	577	11	Q80US9	Q80us9 mus musculu
6	2116.5	68.1	576	13	O42254	O42254 gallus gall
7	1716.5	55.2	594	13	O73932	O73932 xenopus lae
8	1716	55.2	593	13	O57526	O57526 xenopus lae
9	1637	52.6	579	4	O00425	O00425 homo sapien
10	1628	52.3	579	11	Q9CPN8	Q9cpn8 mus musculu
11	1621	52.1	579	11	Q8C2J9	Q8c2j9 mus musculu
12	1612.5	51.8	582	13	Q9PW80	Q9pw80 brachydanio
13	1464	47.1	556	4	Q9Y6M1	Q9y6m1 homo sapien
14	1073	34.5	545	11	Q7TQF9	Q7tqf9 mus musculu
15	858.5	27.6	566	5	Q9VZ69	Q9vz69 drosophila
16	858	27.6	573	5	Q8IR99	Q8ir99 drosophila
17	853.5	27.4	580	5	Q8IGK4	Q8igk4 drosophila
18	459.5	14.8	828	5	Q21605	Q21605 caenorhabdi
19	265	8.5	621	10	Q9C553	Q9c553 arabidopsis
20	257.5	8.3	169	11	Q7TP50	Q7tp50 rattus norv
21	256.5	8.2	398	3	O74919	O74919 schizosacch
22	251	8.1	454	10	Q84MA6	Q84ma6 arabidopsis
23	251	8.1	479	10	Q9XI71	Q9xi71 arabidopsis
24	248	8.0	557	5	Q23487	Q23487 caenorhabdi
25	246.5	7.9	510	5	Q7Z145	Q7z145 caenorhabdi
26	242	7.8	318	13	Q7ZVK5	Q7zvk5 brachydanio
27	240.5	7.7	680	5	P91393	P91393 caenorhabdi
28	240	7.7	672	13	Q8AX85	Q8ax85 xenopus lae
29	238	7.7	542	10	Q8S7G1	Q8s7g1 oryza sativ
30	237.5	7.6	762	10	Q8LNT9	Q8lnt9 oryza sativ
31	236.5	7.6	641	5	Q9BLA0	Q9bla0 caenorhabdi
32	235.5	7.6	632	10	O82762	O82762 arabidopsis
33	234.5	7.5	568	10	Q9LXF5	Q9lxf5 arabidopsis
34	233.5	7.5	313	4	Q96EP6	Q96ep6 homo sapien
35	233	7.5	653	13	Q7ZXS1	Q7zxs1 xenopus lae
36	232	7.5	589	5	Q17935	Q17935 caenorhabdi
37	232	7.5	611	5	Q17936	Q17936 caenorhabdi
38	231	7.4	640	10	Q9ASX3	Q9asx3 arabidopsis
39	231	7.4	644	10	Q9FNK3	Q9fnk3 arabidopsis
40	230	7.4	774	10	Q9LI28	Q9li28 oryza sativ
41	228	7.3	644	10	Q8LDV1	Q8ldv1 arabidopsis
42	226.5	7.3	577	10	Q9SR13	Q9sr13 arabidopsis
43	224	7.2	370	11	Q8BSB0	Q8bsb0 mus musculu
44	222.5	7.2	364	5	Q95SZ9	Q95sz9 drosophila
45	220.5	7.1	353	13	Q9W702	Q9w702 xenopus lae

ALIGNMENTS

RESULT 1
Q8CGX0
ID Q8CGX0 PRELIMINARY; PRT; 577 AA.
AC Q8CGX0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE B-actin zipcode binding protein 1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Eom T., Singer R.H., Bassell G.J.;
RT "Molecular interactions between rZBP1 and b-actin zipcode required for
RT transport of mRNA and stimulation of spine growth.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF541940; AAO16210.1; -.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.

```
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA;  63436 MW;  0647676128FBD1EE CRC64;

Alignment Scores:
Pred. No.:                3.81e-160      Length:          577
Score:                    2221.00        Matches:         440
Percent Similarity:       99.10%         Conservative:    1
Best Local Similarity:    98.88%         Mismatches:      2
Query Match:              71.41%         Indels:          2
DB:                       11             Gaps:            1

US-09-270-437D-5 (1-1708) x Q8CGX0 (1-577)

Qy         32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 ArgGlnAlaIleMetLysLeuAsnGlyHisGlnLeuGluAsnHisAlaLeuLysValSer 152

Qy         92 TACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGC 151
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 TyrIleProAspGluGlnIleAlaGlnGlyProGluAsnGlyArgArgGlyGlyPheGly 172

Qy        152 TCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAG 211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 SerArgGlyGlnProArgGlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGln 192

Qy        212 CAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGC 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GlnValAspIleProLeuArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGly 212

Qy        272 AAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCAT 331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232

Qy        332 AGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGC 391
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGly 252

Qy        392 TGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAA 451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 CysSerSerAlaCysLysMetIleLeuGluIleMetHisLysGluAlaLysAspThrLys 272

Qy        452 ACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATT 511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ThrAlaAspGluValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIle 292

Qy        512 GGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATC 571
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        293 GlyLysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIle 312

Qy        572 TCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCC 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAla 332

Qy        632 ATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAG 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 IleGluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGlu 352

Qy        692 AATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCT 745
              ||||||||||||||||||||      ||||||||||||||||||||||||||||||||||
Db        353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAla 372

Qy        746 GTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGG 805
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 ValGlyLeuPheProAlaSerSerSerAlaValProProProProSerSerValThrGly 392

Qy        806 GCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATC 865
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        393 AlaAlaProTyrGlySerPheMetGlnAlaProGluGlnGluMetValGlnValPheIle 412

Qy        866 CCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCC 925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSer 432

Qy        926 CGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGT 985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg 452

Qy        986 ATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGC 1045
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 MetValValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 472

Qy       1046 AAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATA 1105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        473 LysLeuLysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIle 492

Qy       1106 CGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAG 1165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlu 512

Qy       1166 TTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAAC 1225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        513 LeuGlnAsnLeuThrAlaAlaGluValValValProArgAspGlnThrProAspGluAsn 532

Qy       1226 GACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAG 1285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 AspGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 552

Qy       1286 ATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCC 1345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        553 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerAsnGlnAla 572

Qy       1346 CAGGCACGGAGGAAG 1360
              |||||||||||||||
Db        573 GlnAlaArgArgLys 577

RESULT 2
Q9NZI8
ID   Q9NZI8      PRELIMINARY;      PRT;   577 AA.
AC   Q9NZI8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   mRNA-binding protein CRDBP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ioannidis P., Trangas T., Dimitriadis E., Samiotaki M.,
RA   Panoutsakopoulos G., Kyriazoglou I., Voutzoulias S., Tsiapalis C.M.,
RA   Kittas C., Agnantis N., Pandis N.;
RT   "Ectopic expression of a KH-domain containing protein, highly
RT   homologous to both human IMP-1 and mouse CRD-BP, in benign and
RT   malignant mesenchymal tumors.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF198254; AAF37203.1; -.
DR   HSSP; P11940; 1CVJ.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA;  63456 MW;  0749A060F252D81D CRC64;

Alignment Scores:
```

Pred. No.:	9.15e-160	Length:	577
Score:	2216.00	Matches:	440
Percent Similarity:	98.88%	Conservative:	0
Best Local Similarity:	98.88%	Mismatches:	3
Query Match:	71.25%	Indels:	2
DB:	4	Gaps:	1

US-09-270-437D-5 (1-1708) x Q9NZI8 (1-577)

```
Qy        32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC  91
             |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       133 ArgGlnAlaIleMetLysLeuAsnGlyHisGlnLeuGluAsnHisAlaLeuLysValSer 152

Qy        92 TACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGC 151
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       153 TyrIleProAspGluGlnIleAlaGlnGlyProGluAsnGlyArgArgGlyGlyPheGly 172

Qy       152 TCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAG 211
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       173 SerArgGlyGlnProArgGlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGln 192

Qy       212 CAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGC 271
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       193 GlnValAspIleProLeuArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGly 212

Qy       272 AAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCAT 331
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232

Qy       332 AGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGC 391
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGly 252

Qy       392 TGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAA 451
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       253 CysSerSerAlaCysLysMetIleLeuGluIleMetHisLysGluAlaLysAspThrLys 272

Qy       452 ACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATT 511
             |||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db       273 ThrAlaAspGluValProLeuLysThrLeuAlaHisAsnAsnPheValGlyArgLeuIle 292

Qy       512 GGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATC 571
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       293 GlyLysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIle 312

Qy       572 TCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCC 631
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAla 332

Qy       632 ATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAG 691
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       333 IleGluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGlu 352

Qy       692 AATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCT 745
             |||||||||||||||||||      |||||||||   ||||||||||||||||||||||
Db       353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuThrProGlyLeuAsnLeuAlaAla 372

Qy       746 GTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGG 805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       373 ValGlyLeuPheProAlaSerSerSerAlaValProProProProSerSerValThrGly 392

Qy       806 GCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATC 865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       393 AlaAlaProTyrSerSerPheMetGlnAlaProGluGlnGluMetValGlnValPheIle 412

Qy       866 CCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCC 925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       413 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSer 432

Qy       926 CGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGT 985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       433 ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg 452

Qy       986 ATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGC 1045
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       453 MetValIleIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 472

Qy      1046 AAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATA 1105
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       473 LysLeuLysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIle 492

Qy      1106 CGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAG 1165
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       493 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlu 512

Qy      1166 TTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAAC 1225
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       513 LeuGlnAsnLeuThrAlaAlaGluValValValProArgAspGlnThrProAspGluAsn 532

Qy      1226 GACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAG 1285
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       533 AspGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 552

Qy      1286 ATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCC 1345
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       553 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerAsnGlnAla 572

Qy      1346 CAGGCACGGAGGAAG 1360
             |||||||||||||||
Db       573 GlnAlaArgArgLys 577
```

RESULT 3
O88477
ID O88477 PRELIMINARY; PRT; 577 AA.
AC O88477;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Coding region determinant binding protein.
GN IGF2BP1 OR CRDBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92217743; PubMed=1559612;
RA Bernstein P.L., Herrick D.J., Prokipcak R.D., Ross J.;
RT "Control of c-myc mRNA half-life in vitro by a protein capable of
RT binding to a coding region stability determinant.";
RL Genes Dev. 6:642-654(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94158886; PubMed=8114742;
RA Herrick D.J., Ross J.;
RT "The half-life of c-myc mRNA in growing and serum-stimulated cells:
RT influence of the coding and 3' untranslated regions and role of
RT ribosome translocation.";
RL Mol. Cell. Biol. 14:2119-2128(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94179348; PubMed=8132663;
RA Prokipcak R.D., Herrick D.J., Ross J.;
RT "Purification and properties of a protein that binds to the C-terminal
RT coding region of human c-myc mRNA.";
RL J. Biol. Chem. 269:9261-9269(1994).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=97322234; PubMed=9178888;
RA Leeds P., Kren B.T., Boylan J.M., Betz N.A., Steer C.J.,
RA Gruppuso P.A., Ross J.;
RT "Developmental regulation of CRD-BP, an RNA-binding protein that
RT stabilizes c-myc mRNA in vitro.";
RL Oncogene 14:1279-1286(1997).
RN [5]
RP SEQUENCE FROM N.A.

```
RA   Ross J., Prokipcak R.D., Leeds P., Doyle G.A.R., Betz N.A.,
RA   Fleisig A.J.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF061569; AAC72743.1; -.
DR   EMBL; AK013940; BAB29071.1; -.
DR   HSSP; P11940; 1CVJ.
DR   MGD; MGI:1890357; Igf2bp1.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA;  63450 MW;  EFBB1AF2FF9F0344 CRC64;
```

Alignment Scores:

Pred. No.:	3.71e-159	Length:	577
Score:	2208.00	Matches:	438
Percent Similarity:	98.65%	Conservative:	1
Best Local Similarity:	98.43%	Mismatches:	4
Query Match:	71.00%	Indels:	2
DB:	11	Gaps:	1

US-09-270-437D-5 (1-1708) x O88477 (1-577)

```
Qy        32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
             |||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       133 ArgGlnAlaIleMetLysLeuAsnGlyHisGlnLeuGluAsnHisAlaLeuLysValSer 152

Qy        92 TACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGC 151
             ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db       153 TyrIleProAspGluGlnIleThrGlnGlyProGluAsnGlyArgArgGlyGlyPheGly 172

Qy       152 TCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAG 211
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       173 SerArgGlyGlnProArgGlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGln 192

Qy       212 CAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGC 271
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       193 ProValAspIleProLeuArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGly 212

Qy       272 AAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCAT 331
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232

Qy       332 AGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGC 391
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGly 252

Qy       392 TGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAA 451
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       253 CysSerSerAlaCysLysMetIleLeuGluIleMetHisLysGluAlaLysAspThrLys 272

Qy       452 ACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATT 511
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       273 ThrAlaAspGluValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIle 292

Qy       512 GGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATC 571
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       293 GlyLysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIle 312

Qy       572 TCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCC 631
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAla 332

Qy       632 ATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAG 691
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       333 IleGluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGlu 352

Qy       692 AATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCT 745
             ||||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db       353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAla 372

Qy       746 GTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGG 805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       373 ValGlyLeuPheProAlaSerSerSerAlaValProProProProSerSerValThrGly 392

Qy       806 GCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATC 865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       393 AlaAlaProTyrSerSerPheMetGlnAlaProGluGlnGluMetValGlnValPheIle 412

Qy       866 CCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCC 925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       413 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSer 432

Qy       926 CGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGT 985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       433 ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg 452

Qy       986 ATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGC 1045
             ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       453 MetValValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 472

Qy      1046 AAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATA 1105
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       473 LysLeuLysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIle 492

Qy      1106 CGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAG 1165
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       493 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlu 512

Qy      1166 TTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAAC 1225
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       513 LeuGlnAsnLeuThrAlaAlaGluValValValProArgAspGlnThrProAspGluAsn 532

Qy      1226 GACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAG 1285
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       533 AspGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 552

Qy      1286 ATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCC 1345
             ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       553 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerAsnLeuAla 572

Qy      1346 CAGGCACGGAGGAAG 1360
             |||||||||||||||
Db       573 GlnAlaArgArgLys 577
```

```
RESULT 4
Q8BRH1
ID   Q8BRH1      PRELIMINARY;      PRT;   577 AA.
AC   Q8BRH1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Insulin-like growth factor 2.
GN   IGF2BP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK044850; BAC32119.1; -.
DR   MGD; MGI:1890357; Igf2bp1.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA;  63378 MW;  D75897450841532E CRC64;

Alignment Scores:
Pred. No.:              1.26e-158      Length:          577
Score:                  2201.00        Matches:         437
Percent Similarity:     98.43%         Conservative:    1
Best Local Similarity:  98.20%         Mismatches:      5
Query Match:            70.77%         Indels:          2
DB:                     11             Gaps:            1

US-09-270-437D-5 (1-1708) x Q8BRH1 (1-577)

Qy         32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 ArgGlnAlaIleMetLysLeuAsnGlyHisGlnLeuGluAsnHisAlaLeuLysValSer 152

Qy         92 TACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGC 151
              ||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db        153 TyrIleProAspGluGlnIleThrGlnGlyProGluAsnGlyArgArgGlyGlyPheGly 172

Qy        152 TCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAG 211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 SerArgGlyGlnProArgGlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGln 192

Qy        212 CAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGC 271
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 ProValAspIleProLeuArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGly 212

Qy        272 AAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCAT 331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232

Qy        332 AGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGC 391
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGly 252

Qy        392 TGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAA 451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 CysSerSerAlaCysLysMetIleLeuGluIleMetHisLysGluAlaLysAspThrLys 272

Qy        452 ACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATT 511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ThrAlaAspGluValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIle 292

Qy        512 GGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATC 571
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        293 GlyLysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIle 312

Qy        572 TCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCC 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAla 332

Qy        632 ATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAG 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 IleGluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGlu 352

Qy        692 AATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCT 745
              ||||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAla 372

Qy        746 GTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGG 805
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 ValGlyLeuProAlaSerSerSerAlaValProProProProSerSerValThrGly 392

Qy        806 GCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATC 865
              ||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db        393 AlaAlaProTyrSerSerPheMetGlnAlaProGluGlnGlyMetValGlnValPheIle 412

Qy        866 CCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCC 925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSer 432

Qy        926 CGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGT 985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg 452

Qy        986 ATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGC 1045
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 MetValValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 472

Qy       1046 AAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATA 1105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        473 LysLeuLysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIle 492

Qy       1106 CGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAG 1165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlu 512

Qy       1166 TTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAAC 1225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        513 LeuGlnAsnLeuThrAlaAlaGluValValValProArgAspGlnThrProAspGluAsn 532

Qy       1226 GACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAG 1285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 AspGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 552

Qy       1286 ATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCC 1345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db        553 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerAsnLeuAla 572

Qy       1346 CAGGCACGGAGGAAG 1360
              |||||||||||||||
Db        573 GlnAlaArgArgLys 577

RESULT 5
Q80US9
ID   Q80US9      PRELIMINARY;      PRT;   577 AA.
AC   Q80US9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Igf2bp1 protein.
```

```
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv x 129/Sv-CP; TISSUE=Embryonic stem cells;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv x 129/Sv-CP; TISSUE=Embryonic stem cells;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC051679; AAH51679.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA;  63378 MW;  D439F544257DA3CE CRC64;

Alignment Scores:
Pred. No.:              1.26e-158       Length:         577
Score:                  2201.00         Matches:        437
Percent Similarity:     98.43%          Conservative:   1
Best Local Similarity:  98.20%          Mismatches:     5
Query Match:            70.77%          Indels:         2
DB:                     11              Gaps:           1

US-09-270-437D-5 (1-1708) x Q80US9 (1-577)

Qy         32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 ArgGlnAlaIleMetLysLeuAsnGlyHisGlnLeuGluAsnHisAlaLeuLysValSer 152

Qy         92 TACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGC 151
              ||||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db        153 TyrIleProAspGluGlnIleThrGlnGlyProGluAsnGlyArgArgGlyGlyPheGly 172

Qy        152 TCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAG 211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 SerArgGlyGlnProArgGlnGlySerProValAlaAlaGlyAlaProAlaLysGlnGln 192

Qy        212 CAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGC 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 ProValAspIleProLeuArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGly 212

Qy        272 AAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCAT 331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232

Qy        332 AGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGC 391
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGly 252

Qy        392 TGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAA 451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 CysSerSerAlaCysLysMetIleLeuGluIleMetHisLysGluAlaLysAspThrLys 272

Qy        452 ACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATT 511
              |||||||||    ||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ThrAlaAspGlyValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIle 292

Qy        512 GGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATC 571
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        293 GlyLysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIle 312

Qy        572 TCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCC 631
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAla 332

Qy        632 ATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAG 691
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 IleGluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGlu 352

Qy        692 AATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCT 745
              ||||||||||||||||||||      ||||||||||||||||||||||||||||||||
Db        353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAla 372

Qy        746 GTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGG 805
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 ValGlyLeuPheProAlaSerSerSerAlaValProProProProSerSerValThrGly 392

Qy        806 GCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATC 865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 AlaAlaProTyrSerSerPheMetGlnAlaProGluGlnGluMetValGlnValPheIle 412

Qy        866 CCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCC 925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSer 432

Qy        926 CGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGT 985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg 452

Qy        986 ATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGC 1045
              |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 MetValValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 472

Qy       1046 AAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATA 1105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        473 LysLeuLysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIle 492

Qy       1106 CGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAG 1165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlu 512

Qy       1166 TTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAAC 1225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        513 LeuGlnAsnLeuThrAlaAlaGluValValValProArgAspGlnThrProAspGluAsn 532

Qy       1226 GACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAG 1285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 AspGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 552

Qy       1286 ATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCC 1345
              ||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db        553 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerAsnLeuAla 572

Qy       1346 CAGGCACGGAGGAAG 1360
              |||||||||||||||
```

```
Db        573 GlnAlaArgArgLys 577

RESULT 6
O42254
ID   O42254      PRELIMINARY;      PRT;   576 AA.
AC   O42254;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Zipcode-binding protein.
GN   ZBP1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97220007; PubMed=9121465;
RA   Ross A.F., Oleynikov Y.S., Kislauskis E.H., Taneja K.L., Singer R.H.;
RT   "Characterization of a beta-actin mRNA zipcode-binding protein.";
RL   Mol. Cell. Biol. 17:2158-2165(1997).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF026527; AAB82295.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE   576 AA;  63271 MW;  01AAF2D1D81C8811 CRC64;

Alignment Scores:
Pred. No.:              3.33e-152      Length:          576
Score:                  2116.50        Matches:         419
Percent Similarity:     96.18%         Conservative:    9
Best Local Similarity:  94.16%         Mismatches:      14
Query Match:            68.05%         Indels:          3
DB:                     13             Gaps:            2

US-09-270-437D-5 (1-1708) x O42254 (1-576)

Qy         32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
              |||   |||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        133 ArgGlnAlaIleMetLysLeuAsnGlyHisGlnLeuGluAsnHisValLeuLysValSer 152

Qy         92 TACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGC 151
              ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        153 TyrIleProAspGluGlnSerValGlnGlyProGluAsnGlyArgArgGlyGlyPheGly 172

Qy        152 TCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAG 211
              :::||||||   |||||||||||||||||||||   ||||||||||   |||||||||
Db        173 AlaArgGlyAlaProArgGlnGlySerProValThrAlaGlyAlaProValLysGlnGln 192

Qy        212 CAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGC 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 ProValAspIleProLeuArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGly 212

Qy        272 AAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCAT 331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 LysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHis 232

Qy        332 AGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGC 391
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        233 ArgLysGluAsnAlaGlyAlaAlaGluLysAlaIleSerIleHisSerThrProGluGly 252

Qy        392 TGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAA 451
              ||||||:::|||||||||||||||||||||||||||   ||||||||||||||||||||
Db        253 CysSerAlaAlaCysLysMetIleLeuGluIleMetGlnLysGluAlaLysAspThrLys 272

Qy        452 ACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATT 511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ThrAlaAspGluValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIle 292

Qy        512 GGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATC 571
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        293 GlyLysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIle 312

Qy        572 TCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCC 631
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        313 SerSerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlySer 332

Qy        632 ATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAG 691
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db        333 IleGluAsnCysCysLysAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGlu 352

Qy        692 AATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCT 745
              ||||||||||||||||||||      ||||||||||||||||||||||||||||||||||
Db        353 AsnAspValAlaAlaMetSerLeuGlnSerHisLeuIleProGlyLeuAsnLeuAlaAla 372

Qy        746 GTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGG 805
              ||||||||||||||||||||||||||:::|||||||||||||||||||||||||:::|||
Db        373 ValGlyLeuPheProAlaSerSerAsnAlaValProProProProSerSerValSerGly 392

Qy        806 GCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATC 865
              ||||||||||||||||||||||      ||||||||||   |||   |||||||||
Db        393 AlaAlaProTyrSerSerPheMet---ProProGluGlnGluThrValHisValPheIle 411

Qy        866 CCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCC 925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        412 ProAlaGlnAlaValGlyAlaIleIleGlyLysLysGlyGlnHisIleLysGlnLeuSer 431

Qy        926 CGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGT 985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 ArgPheAlaSerAlaSerIleLysIleAlaProProGluThrProAspSerLysValArg 451

Qy        986 ATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGC 1045
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        452 MetValValIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGly 471

Qy       1046 AAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATA 1105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 LysLeuLysGluGluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIle 491

Qy       1106 CGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAG 1165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 ArgValProAlaSerAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGlu 511

Qy       1166 TTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAAC 1225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        512 LeuGlnAsnLeuThrAlaAlaGluValValValProArgAspGlnThrProAspGluAsn 531

Qy       1226 GACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAG 1285
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 GluGlnValIleValLysIleIleGlyHisPheTyrAlaSerGlnMetAlaGlnArgLys 551

Qy       1286 ATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCC 1345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        552 IleArgAspIleLeuAlaGlnValLysGlnGlnHisGlnLysGlyGlnSerGlyGlnLeu 571

Qy       1346 CAGGCACGGAGGAAG 1360
              |||||||||||||||
Db        572 GlnAlaArgArgLys 576

RESULT 7
O73932
ID   O73932      PRELIMINARY;      PRT;   594 AA.
AC   O73932;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE   VG1 RNA binding protein variant D.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA   Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL   Genes Dev. 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98228351; PubMed=9560341;
RA   Deshler J.O., Highett M.I., Abramson T., Schnapp B.J.;
RT   "A highly conserved RNA-binding protein for cytoplasmic mRNA
RT   localization in vertebrates.";
RL   Curr. Biol. 8:489-496(1998).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF064634; AAC18598.1; -.
DR   EMBL; AF055923; AAC41285.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   594 AA;  65643 MW;  54CEA7BFF0856DD6 CRC64;

Alignment Scores:
Pred. No.:              8.43e-122       Length:         594
Score:                  1716.50         Matches:        343
Percent Similarity:     84.52%          Conservative:   50
Best Local Similarity:  73.76%          Mismatches:     47
Query Match:            55.19%          Indels:         25
DB:                     13              Gaps:           7

US-09-270-437D-5 (1-1708) x O73932 (1-594)

Qy         32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
              |||      :::   |||||||||||:::|||||||||||:::::|||||||:::
Db        133 ArgGlnGlyLeuGluLysLeuAsnGlyTyrGlnLeuGluAsnTyrSerLeuLysValThr 152

Qy         92 TACATCCCCGATGAG-------------------------CAGATAGCACAGGGACCT 124
              |||||||||||||||                         |||:::  |||   |||
Db        153 TyrIleProAspGluMetAlaThrProGlnAlaProSerGlnGlnLeuGlnGlnGlnPro 172

Qy        125 GAG------------AATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAG 172
              :::              |||||||||  |||||||||  ||||||      ||||||
Db        173 GlnGlnGlnHisProGlnGlyArgArg---GlyPheGlyGlnArgGlyProAlaArgGln 191

Qy        173 GGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGG 232
              |||||||||  ||||||     |||      :::   |||   ::::::|||||||||
Db        192 GlySerProGlyAlaAlaAlaArgPro------LysProGlnThrGluValProLeuArg 209

Qy        233 CTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGC 292
              :::|||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        210 MetLeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaThrIleArg 229

Qy        293 AACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCA 352
              |||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        230 AsnIleThrLysGlnThrGlnSerLysIleAspIleHisArgLysGluAsnAlaGlyAla 249

Qy        353 GCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATG 412
              |||||||||   |||::::::|||||||||||||||||||||:::|||||||||:::
Db        250 AlaGluLysProIleThrIleHisSerThrProGluGlyCysSerAlaAlaCysLysIle 269

Qy        413 ATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTG 472
              |||:::|||||||||   |||||||||:::|||||||||      :::|||:::||||||

Db        270 IleMetGluIleMetGlnLysGluAlaGlnAspThrLysPheThrGluGluIleProLeu 289

Qy        473 AAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTG 532
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 LysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeu 309

Qy        533 AAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACC 592
              ||||||:::|||||||||||:::|||||||||||||||||   |||||||||||||||
Db        310 LysLysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnAspLeuThr 329

Qy        593 CTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCC 652
              |||||||||||||||||||||||||||||||||||:::||||||   |||   :::|||
Db        330 LeuTyrAsnProGluArgThrIleThrValLysGlySerIleGluProCysAlaLysAla 349

Qy        653 GAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 712
              |||:::|||||||||||||||:::||||||:::|||||||||||||:::|||||||||:::
Db        350 GluGluGluIleMetLysLysIleArgGluSerTyrGluAsnAspIleAlaAlaMetAsn 369

Qy        713 ------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCA 766
                    :::|||||||||||||||||||||||||   |||:::|||||||||||:::|||
Db        370 LeuGlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProSerSer 389

Qy        767 TCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTT---------ACTGGGGCTGCTCCCTAT 817
              ||||||   :::|||||||||        |||         |||   :::      |||
Db        390 SerSerGlyMetProProProSerValGlyValProSerProThrSerSerThrSerTyr 409

Qy        818 AGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCA 877
                  |||   |||   ||||||   |||   |||   :::|||||||||||   |||
Db        410 ProProPheGlyGlnGlnProGluSerGluThrValHisLeuPheIleProAlaLeuAla 429

Qy        878 GTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGC 937
              |||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        430 ValGlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGly 449

Qy        938 GCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATC 997
              |||||||||||||||||||   |||   |||||:::|||:::|||||||||||||||||
Db        450 AlaSerIleLysIleAlaProAlaGluGlyProAspAlaLysLeuArgMetValIleIle 469

Qy        998 ACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAG 1057
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        470 ThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGlu 489

Qy       1058 GAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCA 1117
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||:::
Db        490 GluAsnPhePheGlyProLysGluGluValLysLeuGluThrHisIleLysValProSer 509

Qy       1118 TCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTG 1177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        510 TyrAlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeu 529

Qy       1178 ACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATC 1237
              |||:::|||||||||||||||||||||||||||||||||||||||||:::|||:::
Db        530 ThrSerAlaGluValValValProArgAspGlnThrProAspGluAsnAspGluValVal 549

Qy       1238 GTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATC 1297
              |||||||||   |||||||||||||||||||::|||||||||||||||:::|||||:|||
Db        550 ValLysIleThrGlyHisPheTyrAlaSerGlnLeuAlaGlnArgLysIleGlnGluIle 569

Qy       1298 CTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGA-----------CAGAGTAACCAGGCC 1345
              ||||||||||:::::::|||   |||:::                |||||   |||
Db        570 LeuAlaGlnValArgArgGlnGlnGlnGlnGlnLysThrValGlnSerGlyGlnPro 589

Qy       1346 CAGGCACGGAGGAAG 1360
              |||   |||||||||
Db        590 GlnProArgArgLys 594

RESULT 8
O57526
ID   O57526      PRELIMINARY;      PRT;   593 AA.
AC   O57526;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE   KH domain-containing transcription factor B3.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92249652; PubMed=1577195;
RA   Pfaff S.L., Taylor W.L.;
RT   "Characterization of a Xenopus oocyte factor that binds to a
RT   developmentally regulated cis-element in the TFIIIA gene.";
RL   Dev. Biol. 151:306-316(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Griffin D., Taylor W.L.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA   Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL   Genes Dev. 0:0-0(1998).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF042353; AAB97457.1; -.
DR   EMBL; AF064633; AAC18597.1; -.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE   593 AA;  65385 MW;  5A5AB4B4A1D55DF7 CRC64;

Alignment Scores:
Pred. No.:              9.2e-122      Length:          593
Score:                  1716.00       Matches:         345
Percent Similarity:     84.27%        Conservative:    46
Best Local Similarity:  74.35%        Mismatches:      49
Query Match:            55.18%        Indels:          24
DB:                     13            Gaps:            7

US-09-270-437D-5 (1-1708) x O57526 (1-593)

Qy         32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
Db        133 ArgGlnGlyLeuGluLysLeuAsnGlyTyrGlnLeuGluAsnTyrSerLeuLysValThr 152

Qy         92 TACATCCCCGATGAG----------------------CAGATAGCACAGGGACCTGAG 127
Db        153 TyrIleProAspGluMetAlaThrProGlnSerProSerGlnGlnLeuGlnGlnProGln 172

Qy        128 ------------AATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGC 175
Db        173 GlnGlnHisProGlnGlyArgArg---GlyPheGlyGlnArgGlyProAlaArgGlnGly 191

Qy        176 TCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTC 235
Db        192 SerProGlyAlaAlaAlaArgProLysProGlnSerGluVal------ProLeuArgMet 209

Qy        236 CTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAAC 295
Db        210 LeuValProThrGlnPheValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsn 229

Qy        296 ATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCT 355
Db        230 IleThrLysGlnThrGlnSerLysIleAspIleHisArgLysGluAsnAlaGlyAlaAla 249

Qy        356 GAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATC 415
Db        250 GluLysProIleThrIleHisSerThrProGluGlyCysSerAlaAlaCysLysIleIle 269

Qy        416 TTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAG 475
Db        270 MetGluIleMetGlnLysGluAlaGlnAspThrLysPheThrGluGluIleProLeuLys 289

Qy        476 ATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAG 535
Db        290 IleLeuAlaHisAsnAsnPheValGlyArgLeuIleGlyLysGluGlyArgAsnLeuLys 309

Qy        536 AAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTT 595
Db        310 LysIleGluGlnAspThrAspThrLysIleThrIleSerProLeuGlnAspLeuThrLeu 329

Qy        596 TACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAG 655
Db        330 TyrAsnProGluArgThrIleThrValLysGlySerIleGluThrCysAlaLysAlaGlu 349

Qy        656 CAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC--- 712
Db        350 GluGluValMetLysLysIleArgGluSerTyrGluAsnAspIleAlaAlaMetAsnLeu 369

Qy        713 ---TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCC 769
Db        370 GlnAlaHisLeuIleProGlyLeuAsnLeuAsnAlaLeuGlyLeuPheProProSerSer 389

Qy        770 AGCGCAGTCCCGCCGCCTCCC---------AGCAGCGTTACTGGGGCTGCTCCCTATAGC 820
Db        390 SerGlyMetProProProSerAlaGlyValSerSerProThrThrSerAlaSerTyrPro 409

Qy        821 TCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTG 880
Db        410 ProPheGlyGlnGlnProGluSerGluThrValHisLeuPheIleProAlaLeuAlaVal 429

Qy        881 GGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCC 940
Db        430 GlyAlaIleIleGlyLysGlnGlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAla 449

Qy        941 TCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACT 1000
Db        450 SerIleLysIleAlaProAlaGluGlyProAspAlaLysLeuArgMetValIleIleThr 469

Qy       1001 GGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAG 1060
Db        470 GlyProProGluAlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGluGlu 489

Qy       1061 AACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCA 1120
Db        490 AsnPhePheGlyProLysGluGluValLysLeuGluAlaHisIleLysValProSerTyr 509

Qy       1121 GCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACG 1180
Db        510 AlaAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThr 529

Qy       1181 GCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTG 1240
Db        530 SerAlaGluValValValProArgAspGlnThrProAspGluAsnAspGlnValValVal 549

Qy       1241 AAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTG 1300
Db        550 LysIleThrGlyHisPheTyrAlaSerGlnLeuAlaGlnArgLysIleGlnGluIleLeu 569

Qy       1301 GCCCAGGTTAAGCAGCAGCATCAGAAG-----------GGACAGAGTAACCAGGCCCAG 1348
Db        570 AlaGlnValArgArgGlnGlnGlnGlnGlnLysThrAlaGlnSerGlyGlnProGln 589

Qy       1349 GCACGGAGGAAG 1360
Db        590 ProArgArgLys 593

RESULT 9
O00425
```

```
ID   O00425      PRELIMINARY;      PRT;   579 AA.
AC   O00425;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative RNA binding protein KOC (KOC).
GN   KOC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mueller-Pillasch F., Lacher U., Wallrapp C., Et AL.;
RL   Oncogene 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Mueller-Pillasch F., Lacher U., Wallrapp C.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; U97188; AAC35208.1; -.
DR   EMBL; U76705; AAD09223.1; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0003723; F:RNA binding; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR   GO; GO:0006412; P:protein biosynthesis; TAS.
DR   GO; GO:0006396; P:RNA processing; TAS.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE   579 AA;  63720 MW;  AE5C3A8EE3C135C5 CRC64;
```

```
Alignment Scores:
Pred. No.:                9.21e-116      Length:          579
Score:                    1637.00        Matches:         336
Percent Similarity:       83.66%         Conservative:    43
Best Local Similarity:    74.17%         Mismatches:      58
Query Match:              52.64%         Indels:          16
DB:                       4              Gaps:            9

US-09-270-437D-5 (1-1708) x O00425 (1-579)
```

```
Qy         32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
Db        133 ArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLysValAla 152

Qy         92 TACATCCCCGATGAGCAGATAGCA---CAGGGACCTGAGAATGGGCGCCGAGGG------ 142
Db        153 TyrIleProAspGluMetAlaAlaGlnGlnAsnProLeuGlnGlnProArgGlyArgArg 172

Qy        143 GGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCC 202
Db        173 GlyLeuGlyGlnArgGlySerSerArgGlnGlySerPro---------GlySerValSer 189

Qy        203 AAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCC 262
Db        190 LysGlnLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheValGlyAla 209

Qy        263 ATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATA 322
Db        210 IleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIle 229

Qy        323 GACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACC 382
Db        230 AspValHisArgLysGluAsnAlaGlyAlaAlaGluLysSerIleThrIleLeuSerThr 249

Qy        383 CCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAG 442
Db        250 ProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGluAlaGln 269

Qy        443 GACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGG 502
Db        270 AspIleLysPheThrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheValGly 289

Qy        503 CGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAA 562
Db        290 ArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspThrLys 309

Qy        563 ATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTG 622
Db        310 IleThrIleSerProLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThrVal 329

Qy        623 AAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAG 682
Db        330 LysGlyAsnValGluThrCysAlaLysAlaGluGluGluIleMetLysLysIleArgGlu 349

Qy        683 GCCTATGAGAATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAAC 736
Db        350 SerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGlyLeuAsn 369

Qy        737 CTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCG---------CCT 787
Db        370 LeuAsnAlaLeuGlyLeuPheProProThrSerGlyMetProProProThrSerGlyPro 389

Qy        788 CCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAG 847
Db        390 ProSerAlaMetThr------ProProTyrProGlnPheGluGlnSer---GluThrGlu 406

Qy        848 ATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAG 907
Db        407 ThrValHisGlnPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGlyGln 426

Qy        908 CACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACA 967
Db        427 HisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluAla 446

Qy        968 CCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCT 1027
Db        447 ProAspAlaLysValArgMetValIleIleThrGlyProProGluAlaGlnPheLysAla 466

Qy       1028 CAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTG 1087
Db        467 GlnGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysGluGluVal 486

Qy       1088 AAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGT 1147
Db        487 LysLeuGluAlaHisIleArgValProSerPheAlaAlaGlyArgValIleGlyLysGly 506

Qy       1148 GGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGAC 1207
Db        507 GlyLysThrValAsnGluLeuGlnAsnLeuSerSerAlaGluValValValProArgAsp 526

Qy       1208 CAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGT 1267
Db        527 GlnThrProAspGluAsnAspGlnValValValLysIleThrGlyHisPheTyrAlaCys 546

Qy       1268 CAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAG---CAGCATCAG 1324
Db        547 GlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGlnGln 566

Qy       1325 AAGGGA---CAGAGTAACCAGGCCCAGGCACGGAGGAAG 1360
Db        567 LysAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
```

```
RESULT 10
Q9CPN8
ID   Q9CPN8      PRELIMINARY;      PRT;   579 AA.
AC   Q9CPN8;
```

```
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   10 days embryo cDNA, RIKEN full-length enriched library,
DE   clone:2610036B18, full insert sequence (Igf2 mRNA-binding protein 3)
DE   (Insulin-like growth factor 2, binding protein 3).
GN   IGF2BP3 OR 2610101N11RIK OR MIMP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Mori H., Sakakibara S., Imai T., Nakamura Y., Iijima T., Suzuki A.,
RA   Yuasa Y., Takeda M., Okano H.;
RT   "Expression of mouse igf2 mRNA-binding protein 3 and its implications
RT   for the developing central nervous system.";
RL   J. Neurosci. Res. 0:0-0(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain, and Olfactory epithelium;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain, and Olfactory epithelium;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AK011689; BAB27779.1; -.
DR   EMBL; AB046173; BAB19755.1; -.
DR   EMBL; BC045138; AAH45138.1; -.
DR   EMBL; BC049082; AAH49082.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   579 AA;  63574 MW;  CABD9A4355B392B7 CRC64;

Alignment Scores:
Pred. No.:              4.45e-115      Length:          579
Score:                  1628.00        Matches:         334
Percent Similarity:     83.92%         Conservative:    47
Best Local Similarity:  73.57%         Mismatches:      55
Query Match:            52.35%         Indels:          18
DB:                     11             Gaps:            10

US-09-270-437D-5 (1-1708) x Q9CPN8 (1-579)

Qy         32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
              |||   ||:::   ||||||||||||   ||||||||||       |||||||||:::
Db        133 ArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLysValAla 152

Qy         92 TACATCCCCGATGAGCAGATAGCACAG---------GGACCTGAG---AATGGGCGCCGA 139
              ||||||||||||||     |||||          |||:::    |||||||||
Db        153 TyrIleProAspGluThrAlaAlaGlnGlnAsnProSerProGlnLeuArgGlyArgArg 172

Qy        140 GGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCA 199
              |||      |||   ||||||      ||||||   ||||||          |||:::
Db        173 GlyPro---GlyGlnArgGlySerSerArgGlnAlaSerPro---------GlySerVal 188

Qy        200 GCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGT 259
              :::||||||:::      |||:::||||||||||||||||||||||||||:::||||||
Db        189 SerLysGlnLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheValGly 208

Qy        260 GCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAG 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 AlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLys 228

Qy        320 ATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCC 379
              ||||||||||||||||||||||  ||||||||||||||:::|||::::::   |||
Db        229 IleAspValHisArgLysGluAsnThrGlyAlaAlaGluLysSerIleThrIleLeuSer 248

Qy        380 ACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCT 439
              ||||||||||||   |||:::||||||||||   ||||||||||||||||||||||||||
Db        249 ThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGluAla 268

Qy        440 AAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTA 499
              :::|||   |||       :::|||:::||||||||||||||||||||||||||||||
Db        269 GlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheVal 288

Qy        500 GGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACA 559
              ||||||||||||||||||||||||||||||||||||||:::||||||||||:::|||
Db        289 GlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspThr 308

Qy        560 AAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACT 619
              |||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||
Db        309 LysIleThrIleSerProLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThr 328

Qy        620 GTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGG 679
              |||||||||:::::::|||   |||   :::||||||:::|||||||||||||||:::|||
Db        329 ValLysGlySerValGluThrCysAlaLysAlaGluGluGluIleMetLysLysIleArg 348

Qy        680 GAGGCCTATGAGAATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTG 733
              |||:::|||||||||||:::|||:::|||:::      :::||||||||||||||||||
Db        349 GluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaHisLeuIleProGlyLeu 368
```

```
Qy        734 AACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCG--------- 784
              ||||||   |||:::|||||||||||||   :::|||         ||||||
Db        369 AsnLeuAsnAlaLeuGlyLeuPheProProThrSerGlyMetProProProThrSerGly 388

Qy        785 CCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAG 844
              |||||||||:::::|||         ||||||      |||   |||:::   |||
Db        389 ProProSerThrLeuThr------ProProTyrProGlnPheGluGlnSer---GluThr 405

Qy        845 GAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGG 904
              |||   |||   :::|||||||||||||   :::|||||||||||||||||||:::|||
Db        406 GluThrValHisLeuPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGly 425

Qy        905 CAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAA 964
              ||||||||||||||||||||||||||||||    ||||||||||||||||||||   |||
Db        426 GlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAlaGlu 445

Qy        965 ACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAG 1024
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 AlaProAspAlaLysValArgMetValIleIleThrGlyProProGluAlaGlnPheLys 465

Qy       1025 GCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAA 1084
              ||||||||||||||||||||||:::|||||||||||||||      ||||||||||||
Db        466 AlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysGluGlu 485

Qy       1085 GTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAA 1144
              ||||||||||||   ||||||||||||||:::   ||||||||||||||||||||||||
Db        486 ValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGlyArgValIleGlyLys 505

Qy       1145 GGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGA 1204
              ||||||||||||||||||||||||||:::|||:::::||||||||||||||||||||||
Db        506 GlyGlyLysThrValAsnGluLeuGlnSerLeuSerSerAlaGluValValValProArg 525

Qy       1205 GACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCC 1264
              ||||||||||||||||||||||||||||||:::|||||||||   ||||||||||||||
Db        526 AspGlnThrProAspGluAsnAspGlnValValValLysIleThrGlyHisPheTyrAla 545

Qy       1265 AGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAG---CAGCAT 1321
              |||:::|||||||||||||||:::::||||||   ||||||||||||   |||
Db        546 CysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGln 565

Qy       1322 CAGAAGGGA---CAGAGTAACCAGGCCCAGGCACGGAGGAAG 1360
              ||||||      ||||||         |||:::|||||||||
Db        566 GlnLysAlaLeuGlnSerGlyProProGlnSerArgArgLys 579

RESULT 11
Q8C2J9
ID   Q8C2J9      PRELIMINARY;      PRT;   579 AA.
AC   Q8C2J9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Insulin-like growth factor 2.
GN   IGF2BP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NOD; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK088465; BAC40370.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   579 AA;  63551 MW;  937E601A95D06B77 CRC64;

Alignment Scores:
Pred. No.:              1.51e-114      Length:         579
Score:                  1621.00        Matches:        333
Percent Similarity:     83.92%         Conservative:   48
Best Local Similarity:  73.35%         Mismatches:     55
Query Match:            52.12%         Indels:         18
DB:                     11             Gaps:           10

US-09-270-437D-5 (1-1708) x Q8C2J9 (1-579)

Qy         32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
              |||   |||:::   |||||||||||   |||||||||||      |||||||||:::
Db        133 ArgGlnAlaLeuAspLysLeuAsnGlyPheGlnLeuGluAsnPheThrLeuLysValAla 152

Qy         92 TACATCCCCGATGAGCAGATAGCACAG---------GGACCTGAG---AATGGGCGCCGA 139
              ||||||||||||||   ||||||          |||:::      |||||||||
Db        153 TyrIleProAspGluThrAlaAlaGlnGlnAsnProSerProGlnLeuArgGlyArgArg 172

Qy        140 GGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCA 199
              |||      |||   ||||||      ||||||   ||||||         |||:::
Db        173 GlyPro---GlyGlnArgGlySerSerArgGlnAlaSerPro---------GlySerVal 188

Qy        200 GCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGT 259
              :::||||||:::      |||:::||||||||||||||||||||||||:::||||||
Db        189 SerLysGlnLysProCysAspLeuProLeuArgLeuLeuValProThrGlnPheValGly 208

Qy        260 GCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAG 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 AlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLys 228

Qy        320 ATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCC 379
              ||||||||||||||||||||||   |||||||||||||:::|||::::::   |||
Db        229 IleAspValHisArgLysGluAsnThrGlyAlaAlaGluLysSerIleThrIleLeuSer 248

Qy        380 ACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCT 439
              ||||||||||||   |||:::|||||||||   ||||||||||||||||||||||||||
Db        249 ThrProGluGlyThrSerAlaAlaCysLysSerIleLeuGluIleMetHisLysGluAla 268

Qy        440 AAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTA 499
              :::|||   |||       :::|||:::||||||||||||||||||||||||||||||
Db        269 GlnAspIleLysPheThrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheVal 288

Qy        500 GGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACA 559
              ||||||||||||||||||||||||||||||||||||:::|||||||||||:::|||
Db        289 GlyArgLeuIleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspThr 308

Qy        560 AAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACT 619
              ||||||||||||||   ||||||:::||||||||||||||||||||||||||||||||
Db        309 LysIleThrIleSerProLeuGlnGluLeuThrLeuTyrAsnProGluArgThrIleThr 328

Qy        620 GTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGG 679
              |||||||||:::::|||   |||   :::||||||:::|||||||||||||||:::|||
Db        329 ValLysGlySerValGluThrCysAlaLysAlaGluGluGluIleMetLysLysIleArg 348

Qy        680 GAGGCCTATGAGAATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTG 733
              |||:::|||||||||||:::|||:::|||:::      ::::::||||||||||||||
Db        349 GluSerTyrGluAsnAspIleAlaSerMetAsnLeuGlnAlaAsnLeuIleProGlyLeu 368

Qy        734 AACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCG--------- 784
              ||||||   |||:::|||||||||||   :::|||         ||||||
Db        369 AsnLeuAsnAlaLeuGlyLeuPheProProThrSerGlyMetProProProThrSerGly 388

Qy        785 CCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAG 844
              |||||||||:::::|||         ||||||      |||   |||:::   |||
Db        389 ProProSerThrLeuThr------ProProTyrProGlnPheGluGlnSer---GluThr 405
```

```
Qy        845 GAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGG 904
Db        406 GluThrValHisLeuPheIleProAlaLeuSerValGlyAlaIleIleGlyLysGlnGly 425

Qy        905 CAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAA 964
Db        426 GlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAlaGlu 445

Qy        965 ACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAG 1024
Db        446 AlaProAspAlaLysValArgMetValIleIleThrGlyProProGluAlaGlnPheLys 465

Qy       1025 GCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAA 1084
Db        466 AlaGlnGlyArgIleTyrGlyLysIleLysGluGluAsnPheValSerProLysGluGlu 485

Qy       1085 GTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAA 1144
Db        486 ValLysLeuGluAlaHisIleArgValProSerPheAlaAlaGlyArgValIleGlyLys 505

Qy       1145 GGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGA 1204
Db        506 GlyGlyLysThrValAsnGluLeuGlnSerLeuSerSerAlaGluValValValProArg 525

Qy       1205 GACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCC 1264
Db        526 AspGlnThrProAspGluAsnAspGlnValValValLysIleThrGlyHisPheTyrAla 545

Qy       1265 AGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAG---CAGCAT 1321
Db        546 CysGlnValAlaGlnArgLysIleGlnGluIleLeuThrGlnValLysGlnHisGlnGln 565

Qy       1322 CAGAAGGGA---CAGAGTAACCAGGCCCAGGCACGGAGGAAG 1360
Db        566 GlnLysAlaLeuGlnSerGlyProProGlnSerArgArgLys 579
```

RESULT 12
Q9PW80
ID Q9PW80 PRELIMINARY; PRT; 582 AA.
AC Q9PW80;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Vg1 RNA binding protein (Decapentaplegic and Vg-related 1, RNA binding
DE protein).
GN DVR1RBP.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Q., Yaniv K., Oberman F., Wolke U., Git A., Fromer M.,
RA Taylor W., Meyer D., Standart N., Raz E., Yisraeli J.K.;
RT "Vg1 RBP intracellular distribution and evolutionarily conserved
RT expression suggest multiple roles during development.";
RL Mech. Dev. 0:0-0(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Body;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; AF161270; AAD45610.1; -.
DR EMBL; BC045873; AAH45873.1; -.
DR ZFIN; ZDB-GENE-000308-1; dvr1rbp.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50084; KH_TYPE_1; 4.
DR PROSITE; PS50102; RRM; 2.
SQ SEQUENCE 582 AA; 63351 MW; 9DAE63200681B306 CRC64;

Alignment Scores:

Pred. No.:	6.71e-114	Length:	582
Score:	1612.50	Matches:	330
Percent Similarity:	81.54%	Conservative:	41
Best Local Similarity:	72.53%	Mismatches:	67
Query Match:	51.85%	Indels:	17
DB:	13	Gaps:	6

US-09-270-437D-5 (1-1708) x Q9PW80 (1-582)

```
Qy         32 CGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCC 91
Db        133 ArgGluAlaMetAspLysLeuAsnGlyPheLeuMetGluAsnTyrAlaLeuLysValSer 152

Qy         92 TACATCCCCGATGAG---CAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTT 148
Db        153 TyrIleProAspGluThrAlaAlaAlaAspAlaProAlaValGlyGlyArgArgGlyPhe 172

Qy        149 GGCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAG 208
Db        173 AsnProArgGlyProProArgGlnGlySerProSerLeuGlyAlaArgPro------Lys 190

Qy        209 CAGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATT 268
Db        191 LeuGlnSerAspValProLeuArgLeuLeuValProThrGlnPheValGlyAlaIleIle 210

Qy        269 GGCAAGGAGGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTG 328
Db        211 GlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrHisSerLysIleAspIle 230

Qy        329 CATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAG 388
Db        231 HisArgLysGluAsnAlaGlyAlaAlaGluLysProIleThrValHisSerThrProGlu 250

Qy        389 GGCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACC 448
Db        251 GlyCysSerSerAlaCysArgAsnIleMetGluIleMetGlnLysGluAlaIleAspThr 270

Qy        449 AAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTC 508
Db        271 LysIleThrGluGluIleProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeu 290
```

```
Qy        509 ATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACC 568
Db        291 IleGlyLysGluGlyArgAsnLeuLysLysIleGluGlnAspThrAspThrLysIleThr 310

Qy        569 ATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGG 628
Db        311 IleSerProLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGly 330

Qy        629 GCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTAT 688
Db        331 ThrLeuAspAlaCysAlaLysAlaGluGluGluIleMetLysLysValArgGluSerTyr 350

Qy        689 GAGAATGATGTGGCTGCCATG------AGCTCTCACCTGATCCCTGGCCTGAACCTGGCT 742
Db        351 GluAsnAspValAlaAlaMetHisLeuGlnSerAsnLeuIleProGlyLeuAsnLeuAsn 370

Qy        743 GCTGTAGGTCTTTTCCCAGCTTCATCCAGC--------------------GCAGTCCCG 781
Db        371 AlaLeuGlyLeuProProGlyAlaAlaSerGlyGlyIleSerProSerValValSerGly 390

Qy        782 CCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAG 841
Db        391 ProProProGlyAlaGlnAlaGly---------TyrGlnSerPheGlyAlaGlnMetGlu 407

Qy        842 CAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAG 901
Db        408 SerGluThrValHisLeuPheIleProAlaLeuAlaValGlyAlaIleIleGlyLysGln 427

Qy        902 GGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCC 961
Db        428 GlyGlnHisIleLysGlnLeuSerArgPheAlaGlyAlaSerIleLysIleAlaProAla 447

Qy        962 GAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATTC 1021
Db        448 AspGlyIleAspAlaLysGlnArgMetValIleIleSerGlyProProGluAlaGlnPhe 467

Qy       1022 AAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAG 1081
Db        468 LysAlaGlnGlyArgIlePheGlyLysLeuLysGluGluAsnPhePheGlyProLysGlu 487

Qy       1082 GAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGC 1141
Db        488 GluValLysLeuGluAlaHisIleLysValProSerPheAlaAlaGlyArgValIleGly 507

Qy       1142 AAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCA 1201
Db        508 LysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrSerAlaGluValValValPro 527

Qy       1202 AGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTAT 1261
Db        528 ArgAspGlnThrProAspGluAsnAspGlnValValValLysIleThrGlyHisPheTyr 547

Qy       1262 GCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCAT 1321
Db        548 AlaSerGlnLeuAlaGlnArgLysIleGlnGluIleIleSerGlnValArgArgGlnGln 567

Qy       1322 CAGAAGGGACAGAGT------AACCAGGCCCAGGCACGGAGGAAG 1360
Db        568 GlnProLysProSerAlaAlaGlyProProValAlaArgArgLys 582
```

```
RESULT 13
Q9Y6M1
ID   Q9Y6M1      PRELIMINARY;      PRT;   556 AA.
AC   Q9Y6M1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hepatocellular carcinoma autoantigen.
GN   P62.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99207072; PubMed=10190901;
RA   Zhang J.Y., Chan E.K., Peng X.X., Tan E.M.;
RT   "A novel cytoplasmic protein with RNA-binding motifs is an autoantigen
RT   in human hepatocellular carcinoma.";
RL   J. Exp. Med. 189:1101-1110(1999).
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AF057352; AAD31596.1; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0003723; F:RNA binding; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR   GO; GO:0009386; P:translational attenuation; TAS.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   556 AA;  61842 MW;  1EDEFB100443DDC4 CRC64;
```

Alignment Scores:

Pred. No.:	1.28e-102	Length:	556
Score:	1464.00	Matches:	292
Percent Similarity:	78.10%	Conservative:	54
Best Local Similarity:	65.91%	Mismatches:	73
Query Match:	47.07%	Indels:	24
DB:	4	Gaps:	5

US-09-270-437D-5 (1-1708) x Q9Y6M1 (1-556)

```
Qy         38 GCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATC 97
Db        136 AlaMetGluLysLeuSerGlyHisGlnPheGluAsnTyrSerPheLysIleSerTyrIle 155

Qy         98 CCCGATGAGCAGATAGCA------CAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGC 151
Db        156 ProAspGluGluValSerSerProSerProProGlnArgAlaGlnArgGlyAspHisSer 175

Qy        152 TCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAG 211
Db        176 SerArgGluGlnGlyHis--------------AlaProGlyGlyThrSerGlnAlaArg 190

Qy        212 CAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGC 271
Db        191 GlnIleAspPheProLeuArgIleLeuValProThrGlnPheValGlyAlaIleIleGly 210

Qy        272 AAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCAT 331
Db        211 LysGluGlyLeuThrIleLysAsnIleThrLysGlnThrGlnSerArgValAspIleHis 230

Qy        332 AGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGC 391
Db        231 ArgLysGluAsnSerGlyAlaAlaGluLysProValThrIleHisAlaThrProGluGly 250

Qy        392 TGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAA 451
Db        251 ThrSerGluAlaCysArgMetIleLeuGluIleMetGlnLysGluAlaAspGluThrLys 270

Qy        452 ACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATT 511
Db        271 LeuAlaGluGluIleProLeuLysIleLeuAlaHisAsnGlyLeuValGlyArgLeuIle 290

Qy        512 GGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATC 571
Db        291 GlyLysGluGlyArgAsnLeuLysLysIleGluHisGluThrGlyThrLysIleThrIle 310

Qy        572 TCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCC 631
Db        311 SerSerLeuGlnAspLeuSerIleTyrAsnProGluArgThrIleThrValLysGlyThr 330
```

```
Qy        632 ATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAG 691
Db        331 ValGluAlaCysAlaSerAlaGluIleGluIleMetLysLysLeuArgGluAlaPheGlu 350

Qy        692 AATGATGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGT 751
Db        351 AsnAspMetLeuAlaValAsnThrHis----------------------------SerGly 361

Qy        752 CTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCT 811
Db        362 TyrPhe---------SerSerLeuTyrProHisHisGln---------------PheGly 373

Qy        812 CCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCC 871
Db        374 ProPheProHisHisHisSerTyrProGluGlnGluIleValAsnLeuPheIleProThr 393

Qy        872 CAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTT 931
Db        394 GlnAlaValGlyAlaIleIleGlyLysLysGlyAlaHisIleLysGlnLeuAlaArgPhe 413

Qy        932 GCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTT 991
Db        414 AlaGlyAlaSerIleLysIleAlaProAlaGluGlyProAspValSerGluArgMetVal 433

Qy        992 ATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTC 1051
Db        434 IleIleThrGlyProProGluAlaGlnPheLysAlaGlnGlyArgIlePheGlyLysLeu 453

Qy       1052 AAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTG 1111
Db        454 LysGluGluAsnPhePheAsnProLysGluGluValLysLeuGluAlaHisIleArgVal 473

Qy       1112 CCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAG 1171
Db        474 ProSerSerThrAlaGlyArgValIleGlyLysGlyGlyLysThrValAsnGluLeuGln 493

Qy       1172 AATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAG 1231
Db        494 AsnLeuThrSerAlaGluValIleValProArgAspGlnThrProAspGluAsnGluGlu 513

Qy       1232 GTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGA 1291
Db        514 ValIleValArgIleIleGlyHisPhePheAlaSerGlnThrAlaGlnArgLysIleArg 533

Qy       1292 GACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCA 1351
Db        534 GluIleValGlnGlnValLysGlnGlnGluGlnLysTyrProGlnGlyValAlaSerGln 553

Qy       1352 CGGAGGAAG 1360
Db        554 ArgSerLys 556
```

```
RESULT 14
Q7TQF9
ID   Q7TQF9      PRELIMINARY;      PRT;   545 AA.
AC   Q7TQF9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC054552; AAH54552.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   545 AA;  59679 MW;  88DD35FF30DF3091 CRC64;
```

```
Alignment Scores:
Pred. No.:              6.59e-73        Length:         545
Score:                  1073.00         Matches:        215
Percent Similarity:     72.58%          Conservative:   47
Best Local Similarity:  59.56%          Mismatches:     59
Query Match:            34.50%          Indels:         40
DB:                     11              Gaps:           5

US-09-270-437D-5 (1-1708) x Q7TQF9 (1-545)

Qy         38 GCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATC 97
Db        135 AlaIleGluLysLeuSerGlyHisGlnPheGluAspTyrSerPheLysIleSerTyrIle 154

Qy         98 CCCGATGAGCAGATAGCA------CAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGC 151
Db        155 ProAspGluGluValSerSerProSerProProHisArgAlaArgGluGlnGlyHisGly 174

Qy        152 TCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAG 211
Db        175 ---------------------------------ProGlySerSerSerGlnAlaArg 182

Qy        212 CAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGC 271
Db        183 GlnIleAspPheProLeuArgIleLeuValProThrGlnPheValGlyAlaIleIleGly 202

Qy        272 AAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCAT 331
Db        203 LysGluGlyLeuThrIleLysAsnIleThrLysGlnThrGlnSerArgValAspIleHis 222

Qy        332 AGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGC 391
Db        223 ArgLysGluAsnSerGlyAlaAlaGluLysProValThrIleHisAlaThrProGluGly 242

Qy        392 TGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAA 451
Db        243 ThrSerGluAlaCysArgMetIleLeuGluIleMetGlnLysGluAlaAspGluThrLys 262

Qy        452 ACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATT 511
Db        263 LeuAlaGluGluValProLeuLysIleLeuAlaHisAsnGlyPheValGlyArgLeuIle 282

Qy        512 GGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATC 571
Db        283 GlyLysGluGlyArgAsnLeuLysLysIleGluHisGluThrGlyThrLysIleThrIle 302

Qy        572 TCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCC 631
```

```
Db        303 SerSerLeuGlnAspLeuSerIleTyrAsnProGluArgThrIleThrValArgGlyThr 322

Qy        632 ATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAG 691
              ||||||   |||      ||||||   ||||||||||||||||:::|||||||||:::|||
Db        323 IleGluAlaCysAlaAsnAlaGluIleGluIleMetLysLysLeuArgGluAlaPheGlu 342

Qy        692 AATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCT 745
              |||||||:::   |||:::::      ::::::|||||||||||||||||||:::|||
Db        343 AsnAspMetLeuAlaValAsnGlnGlnAlaAsnLeuIleProGlyLeuAsnLeuSerAla 362

Qy        746 GTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCT------CCCAGCAGCGTT 799
              :::|||:::|||          |||   :::|||||||||      |||      |||
Db        363 LeuGlyIlePheSerThrGlyLeuSerValLeuProProProAlaGlyProArgGlyVal 382

Qy        800 ACTGGGGCTGCTCCCTATAGCTCCTTTATG---------------------------- 829
                  :::   ||||||      |||
Db        383 ProProSerProProTyrHisProPheAlaThrHisSerGlyTyrPheSerSerLeuTyr 402

Qy        830 ----------------------------------CAGGCTCCCGAGCAGGAGATGGTG 853
                                                 ||||||||||||   |||
Db        403 ProHisHisHisPheGlyProPheProHisHisHisSerTyrProGluGlnGluThrVal 422

Qy        854 CAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATC 913
                :::|||||||||   ||||||||||||||||||||||||||||||||   ||||||
Db        423 SerLeuPheIleProThrGlnAlaValGlyAlaIleIleGlyLysLysGlyAlaHisIle 442

Qy        914 AAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGAC 973
              |||||||||:::|||||||||   ||||||||||||||||||||   |||   ||||||
Db        443 LysGlnLeuAlaArgPheAlaGlyAlaSerIleLysIleAlaProAlaGluGlyProAsp 462

Qy        974 TCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGA 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 ValSerGluArgMetValIleIleThrGlyProProGluAlaGlnPheLysValSerAla 482

Qy       1034 AGA 1036
                 :::
Db        483 Lys 483
```

```
RESULT 15
Q9VZ69
ID   Q9VZ69      PRELIMINARY;      PRT;   566 AA.
AC   Q9VZ69; Q9NGS9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG1691 protein (IGF-II mRNA-binding protein) (SD07045p).
GN   IMP OR CG1691.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Nielsen J., Nielsen F.C., Christiansen J.;
RT   "Cloning and Expression of a Drosophila Homolog of IMP/Vg1-RBP.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR   EMBL; AE003484; AAF47958.2; -.
DR   EMBL; AF241237; AAF63331.1; -.
```

```
DR   EMBL; AY069821; AAL39966.1; -.
DR   FlyBase; FBgn0030235; Imp.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
SQ   SEQUENCE   566 AA;  62129 MW;  B4C74C5C99B0C830 CRC64;
```

Alignment Scores:

Pred. No.:	1.33e-56	Length:	566
Score:	858.50	Matches:	200
Percent Similarity:	57.43%	Conservative:	86
Best Local Similarity:	40.16%	Mismatches:	150
Query Match:	27.60%	Indels:	62
DB:	5	Gaps:	11

US-09-270-437D-5 (1-1708) x Q9VZ69 (1-566)

```
Qy         26 TTACCCCGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAG 85
Db         29 LeuAspArgAlaAlaVal---GlyLeuAsnGlyValGluPheGluGlySerLysLeuHis 47

Qy         86 GTCTCCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGC 145
Db         48 AlaGluGlnLeuAspLysAsnGln---------------------------------- 55

Qy        146 TTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAG 205
Db         56 ------ArgArgSerGlnArgAsnGlnArgAsnProTyr---ProGlyMetProGlyPro 72

Qy        206 CAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATT 265
Db         73 GlyArgGlnAlaAspPheProLeuArgIleLeuValGlnSerGluMetValGlyAlaIle 92

Qy        266 ATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGAC 325
Db         93 IleGlyArgGlnGlySerThrIleArgThrIleThrGlnGlnSerArgAlaArgValAsp 112

Qy        326 GTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCT 385
Db        113 ValHisArgLysGluAsnValGlySerLeuGluLysSerIleThrIleTyrGlyAsnPro 132

Qy        386 GAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGAC 445
Db        133 GluAsnCysThrAsnAlaCysLysArgIleLeuGluValMetGlnGlnGluAlaIleSer 152

Qy        446 ACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGT 505
Db        153 ThrAsnLysGly---GluIleCysLeuLysIleLeuAlaHisAsnAsnLeuIleGlyArg 171

Qy        506 CTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATC 565
Db        172 IleIleGlyLysSerGlyAsnThrIleLysArgIleMetGlnAspThrAspThrLysIle 191

Qy        566 ACCATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAG 625
Db        192 ThrValSerSerIleAsnAspIleAsnSerPheAsnLeuGluArgIleIleThrValLys 211

Qy        626 GGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCC 685
Db        212 GlyLeuIleGluAsnMetSerArgAlaGluAsnGlnIleSerThrLysLeuArgGlnSer 231

Qy        686 TATGAGAATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTG 739
Db        232 TyrGluAsnAspLeuGlnAlaMetAlaProGlnSerLeuMetPheProGlyLeuHisPro 251

Qy        740 GCTGCTGTAGGTCTTTTCCCAGCTTCA------------TCCAGCGCAGTCCCGCCGCCT 787
Db        252 MetAlaMetMetSerThrProGlyAsnGlyMetValPheAsnThrSerMetProPhePro 271

Qy        788 ---------------CCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
```

```
Db        272 SerCysGlnSerPheAlaMetSerLysThrProAlaSerValValProProValPhePro 291

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
Db        292 AsnAspLeuGlnGluThrThrTyrLeuTyrIleProAsnAsnAlaValGlyAlaIleIle 311

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
Db        312 GlyThrArgGlySerHisIleArgSerIleMetArgPheSerAsnAlaSerLeuLysIle 331

Qy        953 GCACCACCCGAAACA------------CCTGACTCCAAAGTTCGTATGGTTATCATCACT 1000
Db        332 AlaProLeuAspAlaAspLysProLeuAspGlnGlnThrGluArgLysValThrIleVal 351

Qy       1001 GGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAG 1060
Db        352 GlyThrProGluGlyGlnTrpLysAlaGlnTyrMetIlePheGluLysMetArgGluGlu 371

Qy       1061 AACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCA 1120
Db        372 GlyPheMetCysGlyThrAspAspValArgLeuThrValGluLeuLeuValAlaSerSer 391

Qy       1121 GCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACG 1180
Db        392 GlnValGlyArgIleIleGlyLysGlyGlyGlnAsnValArgGluLeuGlnArgValThr 411

Qy       1181 GCAGCTGAGGTGGTAGTACCAAGA----------------------------GACCAG 1210
Db        412 GlySerValIleLysLeuProGluHisAlaLeuAlaProProSerGlyGlyAspGluGlu 431

Qy       1211 ACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAG 1270
Db        432 ThrPro--------------------ValHisIleIleGlyLeuPheTyrSerValGln 444

Qy       1271 ATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAG---------------------- 1306
Db        445 SerAlaGlnArgArgIleArgAlaMetMetLeuSerThrAsnProProProIleThrLys 464

Qy       1307 ------------GTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGG 1354
Db        465 LysGlnLysAlaAlaLysGluGlnLeuGlnGlnGlnGlnGlnSerLeuAlaGlyAla-Al 484

Qy       1355 AGGAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCA 1406
Db        484 aSerSerGlySerGlnGlnGlnGlnProGlnSerProSerGlnGlnGlnAla 501
```

Search completed: July 13, 2004, 12:12:37
Job time : 113 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 17:42:43 ; Search time 4592 Seconds
(without alignments)
11107.274 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 1708
Sequence: 1 agggacgctgccgcaccgcc.....atttccttcaggttttaaaa 1708

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1164.8	68.2	2780	11	AK013940	AK013940 Mus muscu
	2	1163.2	68.1	2718	11	AK044850	AK044850 Mus muscu
	3	960.6	56.2	1085	13	BU190334	BU190334 AGENCOURT
	4	912.6	53.4	1042	12	BM928196	BM928196 AGENCOURT
	5	828.4	48.5	1042	12	BM561057	BM561057 AGENCOURT
	6	736.6	43.1	950	13	BQ648383	BQ648383 AGENCOURT
	7	671.8	39.3	831	12	BG115319	BG115319 602316274
c	8	670.4	39.3	1201	13	BX401619	BX401619 BX401619
	9	669	39.2	712	13	BQ225582	BQ225582 AGENCOURT
	10	662	38.8	896	10	BF984962	BF984962 602308484
c	11	660.2	38.7	891	13	BX372273	BX372273 BX372273
	12	653.4	38.3	955	13	BQ651665	BQ651665 AGENCOURT
	13	647.2	37.9	764	12	BM006191	BM006191 603614013
	14	627.6	36.7	983	13	BQ643920	BQ643920 AGENCOURT
	15	627	36.7	4100	11	AK088465	AK088465 Mus muscu
	16	625.4	36.6	2202	11	AK011689	AK011689 Mus muscu
	17	588.4	34.4	978	12	BG115593	BG115593 602317291
	18	582.2	34.1	1173	12	BM467135	BM467135 AGENCOURT
c	19	574.6	33.6	919	13	BX327672	BX327672 BX327672
	20	547.6	32.1	884	13	BU598549	BU598549 AGENCOURT
	21	546	32.0	568	14	CB146278	CB146278 K-EST0201
	22	540	31.6	545	13	BX089485	BX089485 BX089485
	23	537	31.4	1248	29	AY417413	AY417413 Mus muscu
	24	533.8	31.3	1061	13	BQ647561	BQ647561 AGENCOURT
	25	513.8	30.1	953	13	BQ647360	BQ647360 AGENCOURT
	26	513.6	30.1	658	13	BQ256493	BQ256493 NISC_ko03
c	27	513.4	30.1	656	10	AW303505	AW303505 xv19h09.x
c	28	512.4	30.0	550	9	AA978341	AA978341 oq40a04.s
	29	498.4	29.2	643	14	CF894671	CF894671 A0138E10-
c	30	497.2	29.1	517	10	BE550506	BE550506 7a27f01.x
	31	493.8	28.9	1321	29	AY417411	AY417411 Homo sapi
	32	493.4	28.9	1321	29	AY417412	AY417412 Pan trogl
	33	493	28.9	3592	11	AK049196	AK049196 Mus muscu
	34	492	28.8	2964	11	AK044984	AK044984 Mus muscu
c	35	485.4	28.4	761	9	AW003366	AW003366 wq65h07.x
	36	482.6	28.3	3262	11	AK077404	AK077404 Mus muscu
c	37	469	27.5	484	9	AI337147	AI337147 qx83c08.x
	38	459.4	26.9	486	9	AL704124	AL704124 DKFZp686B
c	39	458	26.8	478	9	AA196977	AA196977 zq60h02.s
	40	457	26.8	628	14	CF173461	CF173461 B0924A03-
	41	455.4	26.7	599	14	CF914870	CF914870 B0970D11-
	42	455.4	26.7	628	14	CF172610	CF172610 B0910C03-
	43	453.8	26.6	619	14	CF172952	CF172952 B0916D09-
	44	453.8	26.6	628	14	CF174626	CF174626 B0941G02-
	45	450.4	26.4	540	14	CA561264	CA561264 K0284F02-

ALIGNMENTS

RESULT 1
AK013940
LOCUS AK013940 2780 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus 13 days embryo head cDNA, RIKEN full-length enriched library, clone:3110001D23 product:insulin-like growth factor 2, binding protein 1, full insert sequence.
ACCESSION AK013940
VERSION AK013940.1 GI:12851513
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2780)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.
FEATURES Location/Qualifiers
source 1. .2780
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:3110001D23"
/db_xref="MGI:1904502"
/db_xref="taxon:10090"
/clone="3110001D23"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
CDS 279. .2012
/note="unnamed protein product; insulin-like growth factor 2, binding protein 1 (MGD|MGI:1890357) putative"
/codon_start=1
/protein_id="BAB29071.1"
/db_xref="GI:12851514"
/translation="MNKLYIGNLNESVTPADLEKVFAEHKISYSGQFLVKSGYAFVDC PDEHWAMKAIETFSGKVELQGKRLEIEHSVPKKQRSRKIQIRNIPPQLRWEVLDSLLA QYGTVENCEQVNTESETAVVNVTYSNREQTRQAIMKLNGHQLENHALKVSYIPDEQIT QGPENGRRGGFGSRGQPRQGSPVAAGAPAKQQPVDIPLRLLVPTQYVGAIIGKEGATI RNITKQTQSKIDVHRKENAGAAEKAISVHSTPEGCSSACKMILEIMHKEAKDTKTADE VPLKILAHNNFVGRLIGKEGRNLKKVEQDTETKITISSLQDLTLYNPERTITVKGAIE NCCRAEQEIMKKVREAYENDVAAMSLQSHLIPGLNLAAVGLFPASSSAVPPPPSSVTG AAPYSSFMQAPEQEMVQVFIPAQAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSK VRMVVITGPPEAQFKAQGRIYGKLKEENFFGPKEEVKLETHIRVPASAAGRVIGKGGK TVNELQNLTAAEVVVPRDQTPDENDQVIVKIIGHFYASQMAQRKIRDILAQVKQQHQK GQSNLAQARRK"

ORIGIN

Query Match 68.2%; Score 1164.8; DB 11; Length 2780;
Best Local Similarity 82.7%; Pred. No. 8.6e-289;
Matches 1412; Conservative 0; Mismatches 262; Indels 33; Gaps 6;

```
Qy        30 CCCGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 89
             || ||  ||| |||||||||| |||||||| ||  ||||||||||||||||||||||||
Db       673 CCAGGCAAGCTATCATGAAGCTAAATGGCCATCAACTGGAGAACCATGCCCTGAAGGTCT 732

Qy        90 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 149
             ||||||| || |||||||||||||  | || || |||||||||||||||| || || |||||||
Db       733 CCTACATACCTGATGAGCAGATAACGCAAGGTCCTGAGAATGGGCGTCGTGGAGGCTTTG 792

Qy       150 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 209
             | |||||||| |||||||| || || || || |||||||| ||||| ||||||||||||||
Db       793 GGTCTCGGGGCCAGCCCCGGCAAGGGTCGCCCGTGGCAGCAGGGGCTCCAGCCAAGCAGC 852

Qy       210 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 269
             ||| |||||||||||| || |||||||||||||| || ||||||| || || || ||||
Db       853 AGCCAGTGGACATCCCTCTCCGGCTCCTGGTGCCTACGCAGTATGTAGGCGCTATCATTG 912

Qy       270 GCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 329
             |||||||||| |||||||||| |||||||||||||||||| |||||||| ||||||||||
Db       913 GCAAGGAGGGTGCCACCATCCGAAACATCACAAAACAGACGCAGTCCAAAATAGACGTGC 972

Qy       330 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGG 389
             |||||||||||| || || || || || |||||||| |||| || |||||||||||| |
Db       973 ATAGGAAGGAGAATGCGGGCGCTGCGGAGAAGGCCATCAGCGTGCATTCAACCCCTGAAG 1032

Qy       390 GCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCA 449
             |||||||||||| || |||||||||||||||||||||||| || |||| ||||||||||
Db      1033 GCTGCTCCTCCGCGTGCAAGATGATCTTGGAGATTATGCACAAGGAGGCAAAGGACACCA 1092

Qy       450 AAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCA 509
             ||||||| || || |||||||||||||||||| |||| || ||||| |||| |||||||
Db      1093 AAACGGCAGATGAAGTTCCCCTGAAGATCCTGGCTCATAACAACTTCGTCGGGCGACTCA 1152

Qy       510 TTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCA 569
             |||||||||||| |||||||||||||||||| |||| || || |||| || |||||||
Db      1153 TTGGCAAGGAAGGGCGGAACCTGAAGAAGGTGGAGCAGGACACAGAGACGAAGATCACCA 1212

Qy       570 TCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGG 629
             |||| ||| | || |||| || || || |||||||||||||||||||||||||||||| |
Db      1213 TCTCATCGCTCCAGGACCTCACGCTCTATAACCCTGAGAGGACCATCACTGTGAAGGGCG 1272

Qy       630 CCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATG 689
             |||| |||| |||||||||||||||||||| || |||||||||||||||| |||| || |
Db      1273 CCATTGAGAACTGTTGCAGGGCCGAGCAGGAGATCATGAAGAAAGTTCGAGAGGCTTACG 1332

Qy       690 AGAATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTG 743
             |||| || |||| |||||||||       || ||||| |||||||| || ||||||||||
Db      1333 AGAACGACGTGGCCGCCATGAGCTTGCAGTCCCACCTCATCCCTGGGCTTAACCTGGCTG 1392

Qy       744 CTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTG 803
             ||||||||| |||||||||||||||||||| |||| || |||||||||| || || |
Db      1393 CTGTAGGTCTCTTCCCAGCTTCATCCAGCGCTGTCCCTCCTCCTCCCAGCAGTGTCACCG 1452
```

```
Qy        804 GGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTA 863
              |||||||||||||||||||||| |||||||||| |||||||||||||| || |||||| |
Db       1453 GGGCTGCTCCCTATAGCTCCTTCATGCAGGCTCCGGAGCAGGAGATGGTACAAGTGTTCA 1512

Qy        864 TCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCT 923
              |||||||||||| |||||||||||| |||||||||| |||||||||||||| ||||
Db       1513 TCCCCGCCCAGGCTGTGGGCGCCATCATTGGCAAGAAGGGCCAGCACATCAAACAACTCT 1572

Qy        924 CCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTC 983
              |||| || |||||||||||||||||||||||||||||| |||||||||||||||||||||
Db       1573 CCCGCTTCGCCAGCGCCTCCATCAAGATTGCACCACCAGAAACACCTGACTCCAAAGTTC 1632

Qy        984 GTATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATG 1043
              | |||||  |||||||||||| |||||||| || |||||||| ||||||||||||||||
Db       1633 GAATGGTCGTCATCACTGGACCCCCAGAGGCTCAGTTCAAGGCCCAGGGAAGAATCTATG 1692

Qy       1044 GCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACA 1103
              ||||||| || || ||||| |||||||||||||||||||||||| ||||| ||||||||||
Db       1693 GCAAACTAAAAGAAGAGAATTTCTTTGGTCCCAAGGAGGAAGTAAAGCTAGAGACCCACA 1752

Qy       1104 TACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACG 1163
              |||| || || || |||||||| ||||| ||||| |||||||| || |||||||||||| |
Db       1753 TACGGGTTCCGGCTTCAGCAGCCGGCCGTGTCATCGGCAAAGGCGGCAAAACGGTGAATG 1812

Qy       1164 AGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGA 1223
              || ||||||| ||||| |||||||||||||||||||||||||| |||||||||||||||||
Db       1813 AGCTGCAGAACTTGACCGCAGCTGAGGTGGTAGTGCCAAGAGACCAGACCCCGGATGAGA 1872

Qy       1224 ACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGA 1283
              ||||||| ||||| || || |||||||||||||||||||||| |||||||||||| ||||
Db       1873 ACGACCAAGTCATTGTTAAGATCATCGGACATTTCTATGCCAGCCAGATGGCTCAGCGGA 1932

Qy       1284 AGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGG 1343
              |||||||||||||||||| || |||||||| ||||| |||||||||||||| |||| ||
Db       1933 AGATCCGAGACATCCTGGCTCAAGTTAAGCAACAGCACCAGAAGGGACAGAGCAACCTGG 1992

Qy       1344 CCCAGGCACGGAGGAAGTGA-CCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACG 1402
              ||||||| |||||||||||| || |||||  |||||||| | |  |||| ||  |  | |
Db       1993 CCCAGGCTCGGAGGAAGTGACCCCGCCCCCTCCTGTCCCATTGGCTCCAAGATCAGCAGG 2052

Qy       1403 GGCAGAA--------------------ATCGAGAGTGTGCTCTCCCCGGCAGGCCTGAG 1441
                | | |                    | |   ||||||||| ||| ||||||||||
Db       2053 AGGAACACAGAACTGGAGGGGCGGGTGGAGGGCCGGTGTGCTCTTCCCAGCAGGCCTGAG 2112

Qy       1442 AATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATTG 1501
              |||||||||||||| ||| ||  |||||| ||||| |||||||||||| ||||   |||
Db       2113 AATGAGTGGGAATCAGGG-CATTTGGGCCTGGCTGGAGATCAGGTTTGCACACTGTCTTG 2171

Qy       1502 AGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGCC 1561
              ||||  |||||||||||||||| |||||||| ||  ||||||       ||   | | |
Db       2172 AGAACAATGTTCCAGTGAGGAATCCTGATCTCTCGCCCCCAA--TTGAGCCAGCTGGCCA 2229

Qy       1562 CAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGGA 1621
              || | |                  | |||  ||||    |  ||||||||||||||
Db       2230 CAGCCCACCCCTTGGAATATCACCATTGCAATCATAGCTTGGGTTGCTTTTAAACGTGGA 2289

Qy       1622 TTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGAA 1681
              |||| |  |||| ||||||| | |||  | |  ||    ||     |||||||
Db       2290 TTGTCT--TGAAGTTCTCCAGCCTCCATGGAAGGATGGGTCAGATCCCAGTGGGGAAGAG 2347

Qy       1682 AAATAAAATTTCCTTCAGGTTTTAAAA 1708
              |||||||||||||||||||||||| ||
Db       2348 AAATAAAATTTCCTTCAGGTTTTATAA 2374
```

RESULT 2
AK044850

```
LOCUS       AK044850               2718 bp    mRNA    linear   HTC 20-SEP-2003
DEFINITION  Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130008A04 product:insulin-like growth
            factor 2, binding protein 1, full insert sequence.
ACCESSION   AK044850
VERSION     AK044850.1  GI:26336870
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2718)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. Tomohiro Kono (Department of Animal
            Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
            Kanagawa Prefecture, Japan) whose assistance we gratefully
```

```
          acknowledge.
          Please visit our web site for further details.
          URL:http://genome.gsc.riken.go.jp/
          URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .2718
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B130008A04"
                     /db_xref="MGI:2410206"
                     /db_xref="taxon:10090"
                     /clone="B130008A04"
                     /tissue_type="parthenogenote"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="9.5 days embryo"
     CDS             217. .1950
                     /note="unnamed protein product; insulin-like growth factor
                     2, binding protein 1 (MGD|MGI:1890357)
                     putative"
                     /codon_start=1
                     /protein_id="BAC32119.1"
                     /db_xref="GI:26336871"
                     /translation="MNKLYIGNLNESVTPADLEKVFAEHKISYSGQFLVKSGYAFVDC
                     PDEHWAMKAIETFSGKVELQGKRLEIEHSVPKKQRSRKIQIRNIPPQLRWEVLDSLLA
                     QYGTVENCEQVNTESETAVVNVTYSNREQTRQAIMKLNGHQLENHALKVSYIPDEQIT
                     QGPENGRRGGFGSRGQPRQGSPVAAGAPAKQQPVDIPLRLLVPTQYVGAIIGKEGATI
                     RNITKQTQSKIDVHRKENAGAAEKAISVHSTPEGCSSACKMILEIMHKEAKDTKTADE
                     VPLKILAHNNFVGRLIGKEGRNLKKVEQDTETKITISSLQDLTLYNPERTITVKGAIE
                     NCCRAEQEIMKKVREAYENDVAAMSLQSHLIPGLNLAAVGLFPASSSAVPPPPSSVTG
                     AAPYSSFMQAPEQGMVQVFIPAQAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSK
                     VRMVVITGPPEAQFKAQGRIYGKLKEENFFGPKEEVKLETHIRVPASAAGRVIGKGGK
                     TVNELQNLTAAEVVVPRDQTPDENDQVIVKIIGHFYASQMAQRKIRDILAQVKQQHQK
                     GQSNLAQARRK"
ORIGIN

  Query Match           68.1%;  Score 1163.2;  DB 11;  Length 2718;
  Best Local Similarity  82.7%;  Pred. No. 2.2e-288;
  Matches 1411;  Conservative    0;  Mismatches  263;  Indels   33;  Gaps    6;

Qy         30 CCCGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 89
Db        611 CCAGGCAAGCTATCATGAAGCTAAATGGCCATCAACTGGAGAACCATGCCCTGAAGGTCT 670

Qy         90 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 149
Db        671 CCTACATACCTGATGAGCAGATAACGCAAGGTCCTGAGAATGGGCGTCGTGGAGGCTTTG 730

Qy        150 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 209
Db        731 GGTCTCGGGGCCAGCCCCGGCAAGGGTCGCCCGTGGCAGCAGGGGCTCCAGCCAAGCAGC 790

Qy        210 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 269
Db        791 AGCCAGTGGACATCCCTCTCCGGCTCCTGGTGCCTACGCAGTATGTAGGCGCTATCATTG 850

Qy        270 GCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 329
Db        851 GCAAGGAGGGTGCCACCATCCGAAACATCACAAAACAGACGCAGTCCAAAATAGACGTGC 910

Qy        330 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGG 389
Db        911 ATAGGAAGGAGAATGCGGGCGCTGCGGAGAAGGCCATCAGCGTGCATTCAACCCCTGAAG 970

Qy        390 GCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCA 449
Db        971 GCTGCTCCTCCGCGTGCAAGATGATCTTGGAGATTATGCACAAGGAGGCAAAGGACACCA 1030

Qy        450 AAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCA 509
Db       1031 AAACGGCAGATGAAGTTCCCCTGAAGATCCTGGCTCATAACAACTTCGTCGGGCGACTCA 1090

Qy        510 TTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCA 569

Db       1091 TTGGCAAGGAAGGGCGGAACCTGAAGAAGGTGGAGCAGGACACAGAGACGAAGATCACCA 1150

Qy        570 TCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGG 629
Db       1151 TCTCATCGCTCCAGGACCTCACGCTCTATAACCCTGAGAGGACCATCACTGTGAAGGGCG 1210

Qy        630 CCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATG 689
Db       1211 CCATTGAGAACTGTTGCAGGGCCGAGCAGGAGATCATGAAGAAAGTTCGAGAGGCTTACG 1270

Qy        690 AGAATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTG 743
Db       1271 AGAACGACGTGGCCGCCATGAGCTTGCAGTCCCACCTCATCCCTGGGCTTAACCTGGCTG 1330

Qy        744 CTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTG 803
Db       1331 CTGTAGGTCTCTTCCCAGCTTCATCCAGCGCTGTCCCTCCTCCTCCCAGCAGTGTCACCG 1390

Qy        804 GGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTA 863
Db       1391 GGGCTGCTCCCTATAGCTCCTTCATGCAGGCTCCGGAGCAGGGGATGGTACAAGTGTTCA 1450

Qy        864 TCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCT 923
Db       1451 TCCCCGCCCAGGCTGTGGGCGCCATCATTGGCAAGAAGGGCCAGCACATCAAACAACTCT 1510

Qy        924 CCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTC 983
Db       1511 CCCGCTTCGCCAGCGCCTCCATCAAGATTGCACCACCAGAAACACCTGACTCCAAAGTTC 1570

Qy        984 GTATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATG 1043
Db       1571 GAATGGTCGTCATCACTGGACCCCCAGAGGCTCAGTTCAAGGCCCAGGGAAGAATCTATG 1630

Qy       1044 GCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACA 1103
Db       1631 GCAAACTAAAAGAAGAGAATTTCTTTGGTCCCAAGGAGGAAGTAAAGCTAGAGACCCACA 1690

Qy       1104 TACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACG 1163
Db       1691 TACGGGTTCCGGCTTCAGCAGCCGGCCGTGTCATCGGCAAAGGCGGCAAAACGGTGAATG 1750

Qy       1164 AGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGA 1223
Db       1751 AGCTGCAGAACTTGACCGCAGCTGAGGTGGTAGTGCCAAGAGACCAGACCCCGGATGAGA 1810

Qy       1224 ACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGA 1283
Db       1811 ACGACCAAGTCATTGTTAAGATCATCGGACATTTCTATGCCAGCCAGATGGCTCAGCGGA 1870

Qy       1284 AGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGG 1343
Db       1871 AGATCCGAGACATCCTGGCTCAAGTTAAGCAACAGCACCAGAAGGGACAGAGCAACCTGG 1930

Qy       1344 CCCAGGCACGGAGGAAGTGA-CCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACG 1402
Db       1931 CCCAGGCTCGGAGGAAGTGACCCCGCCCCCTCCTGTCCCATTGGCTCCAAGATCAGCAGG 1990

Qy       1403 GGCAGAA----------------------ATCGAGAGTGTGCTCTCCCCGGCAGGCCTGAG 1441
Db       1991 AGGAACACAGAACTGGAGGGGCGGGTGGAGGGCCGGTGTGCTCTTCCCAGCAGGCCTGAG 2050

Qy       1442 AATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATTG 1501
Db       2051 AATGAGTGGGAATCAGGG-CATTTGGGCCTGGCTGGAGATCAGGTTTGCACACTGTCTTG 2109

Qy       1502 AGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGCC 1561
Db       2110 AGAACAATGTTCCAGTGAGGAATCCTGATCTCTCGCCCCCAA--TTGAGCCAGCTGGCCA 2167

Qy       1562 CAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGGA 1621
```

```
Db       2168 CAGCCCACCCCTTGGAATATCACCATTGCAATCATAGCTTGGGTTGCTTTTAAACGTGGA 2227

Qy       1622 TTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGAA 1681

Db       2228 TTGTCT--TGAAGTTCTCCAGCCTCCATGGAAGGATGGGTCAGATCCCAGTGGGGAAGAG 2285

Qy       1682 AAATAAAATTTCCTTCAGGTTTTAAAA 1708

Db       2286 AAATAAAATTTCCTTCAGGTTTTATAA 2312


RESULT 3
BU190334
LOCUS       BU190334               1085 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7186901 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6001834
            5', mRNA sequence.
ACCESSION   BU190334
VERSION     BU190334.1  GI:22704318
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1085)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2269  row: b  column: 11
            High quality sequence stop: 671.
FEATURES             Location/Qualifiers
     source          1. .1085
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6001834"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match           56.2%;  Score 960.6;  DB 13;  Length 1085;
  Best Local Similarity   95.0%;  Pred. No. 2.5e-236;
  Matches 1026;  Conservative    0;  Mismatches   45;  Indels    9;  Gaps    3;

Qy         30 CCCGGGGAGCCATCATGAAGCTGAAT-GGCCACCAGTTGGAGAACCATGCCCTGAAGGTC 88

Db          5 CCAGGCAAGCCATCATGAAGCTGAATGGGCCACCAGTTGGAGAACCATGCCCTGAAGGTC 64

Qy         89 TCCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTT 148

Db         65 TCCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTT 124

Qy        149 GGCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAG 208

Db        125 GGCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAG 184

Qy        209 CAGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATT 268

Db        185 CAGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATT 244

Qy        269 GGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTG 328

Db        245 GGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTG 304

Qy        329 CATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAG 388

Db        305 CATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAG 364

Qy        389 GGCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACC 448

Db        365 GGCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACC 424

Qy        449 AAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTC 508

Db        425 AAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTC 484

Qy        509 ATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACC 568

Db        485 ATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACC 544

Qy        569 ATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGG 628

Db        545 ATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGG 604

Qy        629 GCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTAT 688

Db        605 GCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTAT 664

Qy        689 GAGAATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCT 742

Db        665 GAGAATGATGTGGCTGCCATGAGCCTGCAGTCTCACCTGATCCCTGGCCTGAACCTGGCT 724

Qy        743 GCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACT 802

Db        725 GCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACT 784

Qy        803 GGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTT 862

Db        785 GGGGCTGCTCCCTATAGCTCCTTTATGCANGCTCCCGAGCAGGAGATGGTGCAGGTGTTT 844

Qy        863 ATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTC 922

Db        845 ATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTC 904

Qy        923 TCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTT 982

Db        905 TCCCGGTTTGCCCAGGCCTTCCTCAAGATTGCACACCCGAAACACCCGGACTCCAAAGTT 964

Qy        983 CGTATGGTTATCATCAC--TGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCT 1040

Db        965 CGAATGGTTATCCATACCTGGAACGCCCAAAGGCCCAATTCCAGGCTCAGGGGAAAATCT 1024

Qy       1041 ATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCC 1100

Db       1025 ATGGGCAATTCAGGGAGGAAAACTCCTTTGGCCCCAGGGAGGAATGAACCCTGGGAAACC 1084


RESULT 4
BM928196
LOCUS       BM928196               1042 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6715416 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5797961
            5', mRNA sequence.
ACCESSION   BM928196
VERSION     BM928196.1  GI:19378575
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1042)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2025  row: c  column: 18
            High quality sequence stop: 662.
FEATURES             Location/Qualifiers
     source          1. .1042
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5797961"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match           53.4%;  Score 912.6;  DB 12;  Length 1042;
  Best Local Similarity   97.5%;  Pred. No. 5.9e-224;
  Matches  981;  Conservative    0;  Mismatches   15;  Indels   10;  Gaps    5;

Qy         30 CCCGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 89
              ||  ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 CCAGGCAAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 64

Qy         90 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 124

Qy        150 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 184

Qy        210 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 244

Qy        270 GCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 304

Qy        330 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGG 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGG 364

Qy        390 GCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCA 424

Qy        450 AAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCA 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 AAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCA 484

Qy        510 TTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCA 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 TTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCA 544

Qy        570 TCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 TCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGG 604

Qy        630 CCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATG 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 CCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATG 664

Qy        690 AGAATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTG 743
              |||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        665 AGAATGATGTGGCTGCCATGAGCCTGCAGTCTCACCTGATCCCTGGCCTGAACCTGGCTG 724

Qy        744 CTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTG 803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        725 CTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTG 784

Qy        804 GGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTA 863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        785 GGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCANGAGATGGTGCAGGTGTTTA 844

Qy        864 T-CCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTC 922
              | ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        845 TCCCCCGCCCAGGCAGTGGGCGCCATCATCGGGCAGAAGGGGCAGCACATCAAACAGCTC 904

Qy        923 TCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCC-AAAGT 981
              |||| | ||||||||||||| ||||||||||||||||||||||||||||||| |||||
Db        905 TCCCCGGTTGCCAGCGCCTTCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAAGT 964

Qy        982 TCGTAT-GGTTATCATCAC-TGGACCGCCAGAGGCCCAATTCAAGG 1025
              |||||| |||||||||| | ||||||| |||| ||||||||  | |
Db        965 TCGTATGGGTTATCATCCCTTGGACCGGCAGAAGCCCAATTTCAAG 1010


RESULT 5
BM561057
LOCUS       BM561057                1042 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6566228 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5550887
            5', mRNA sequence.
ACCESSION   BM561057
VERSION     BM561057.1  GI:18805998
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1042)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12263  row: l  column: 24
            High quality sequence stop: 682.
FEATURES             Location/Qualifiers
     source          1. .1042
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5550887"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
```

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."

ORIGIN

Query Match 48.5%; Score 828.4; DB 12; Length 1042;
Best Local Similarity 96.8%; Pred. No. 3.1e-202;
Matches 920; Conservative 0; Mismatches 18; Indels 12; Gaps 7;

```
Qy        438 CTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTG 497
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 CTCAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTG 61

Qy        498 TAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGA 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGA 121

Qy        558 CAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCA 617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCA 181

Qy        618 CTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTC 677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 CTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTC 241

Qy        678 GGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCC 731
              |||||||||||||||||||||||||||||||||      ||||||||||||||||||||
Db        242 GGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCCTGCAGTCTCACCTGATCCCTGGCC 301

Qy        732 TGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCA 791
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCA 361

Qy        792 GCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGG 851
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGG 421

Qy        852 TGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACA 911
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACA 481

Qy        912 TCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTG 971
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 TCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTG 541

Qy        972 ACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGG 1031
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 ACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGG 601

Qy       1032 GAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGC 1091
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 GAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGC 661

Qy       1092 TGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAA 1151
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 TGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAA 721

Qy       1152 AAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGA 1211
              |||||||||||||||||||||||||||||||||    |||||||||||||||| ||||
Db        722 AAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTNGAGTGGTAGTACCAAGAGAACAGA 781

Qy       1212 CCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATC-GGACATTTCTATGCCAGTCAG 1270
              ||||||||||||    |  ||||||||||||||||| ||||||||||||||||||||||
Db        782 CCCCTGATGAGAAACGACCAGTCATCGTGAAAATCATCGGGACATTTCTATGCCAGTCAG 841

Qy       1271 ATGGCTCAAC-GGAAGATCCGAGACATCCT-GGCCCAGGTTAAGCAGCAGCAT-CAGAAG 1327
              |||||||  ||  ||||||||||||||||| |||||||||||||||||||||| || |||
Db        842 ATGGCTCCACGGGAAGATCCGAGACATCCTGGGCCCAGGTTAAGCAGCAGCATCCAAAAG 901

Qy       1328 GGACAGA-GTAACCAGGCCCAGGCAC-GGAGGAAGTGACCAGCCCCTCCC 1375
              |||||||  ||||||||||||  ||||| ||||||||||||||| |||  ||
Db        902 GGACAGAGGTAACCAGGCCAGGGCACGGGAGGAAGTGACCAGNCCCCTCC 951
```

RESULT 6
BQ648383
LOCUS BQ648383 950 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8342265 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6268202 5', mRNA sequence.
ACCESSION BQ648383
VERSION BQ648383.1 GI:21772555
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 950)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2439 row: m column: 03
High quality sequence stop: 599.
FEATURES Location/Qualifiers
source 1. .950
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6268202"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 43.1%; Score 736.6; DB 13; Length 950;
Best Local Similarity 93.9%; Pred. No. 1.4e-178;
Matches 857; Conservative 0; Mismatches 39; Indels 17; Gaps 8;

```
Qy         30 CCCGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 89
              || ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 CCAGGCAAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 64

Qy         90 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 124

Qy        150 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 184

Qy        210 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 244

Qy        270 GCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 304

Qy        330 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGG 389
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGG 364

Qy        390 GCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCA 449
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCA 424

Qy        450 AAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCA 509
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 AAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCA 484

Qy        510 TTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCA 569
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 TTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCA 544

Qy        570 TCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACT-GTGAAGGGG 628
                 ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        545 TCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGGTGAAGGGG 604

Qy        629 GCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTAT 688
                 ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db        605 GCCATCGAGAATTGTTGCAGGGCCGAGCAGG-AATAATGAAGAAAGTTCGGGAGGCCTAT 663

Qy        689 GAGAATGATGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAAACCT-GGC 741
                 ||||||||||||||||||||||      |||||||||||||||||||||| || |||
Db        664 GAGAATGATGTGGCTGCCATGAGCCTGCAGTCTCACCTGATCCCTGGCCTGAAACTGGGC 723

Qy        742 TGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGC-GTTA 800
                 ||||||||||||||| |||||||||| |||||||||||||||||  |||||||| || |
Db        724 TGCTGTAGGTCTTTTTCCAGCTTCATTCAGCGCAGTCCCGCCGCCCTCCAGCAGCGGTAA 783

Qy        801 CTGGGGCTGCT-CCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTG--CAGG 857
                 |||||||||| |||||||||||||||| ||| |||||| | | ||| | || |    |||
Db        784 CTGGGGCTGCTCCCCTATAGCTCCTTTAAGCAAGCTCCCCAACCGGAAAAGGGGGCAAGG 843

Qy        858 TGTTTATCCCCGCCCAGGCAGTGGGCGC----CATCATCGGCAAGAAGGGGCAGCACATC 913
                 ||| | |||| |||| | ||  || ||     |      |||||| | |||| |||||  ||
Db        844 GGTTAACCCCCCCCCAAGGAGGTGGGGCGCCACTCATCCGGCAAAAGGGGGTAGCACCTC 903

Qy        914 AAACAGCTCTCCC 926
                 |||| ||  ||||
Db        904 AAACCGCCTTCCC 916


RESULT 7
BG115319
LOCUS       BG115319                831 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  602316274F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4416354 5',
            mRNA sequence.
ACCESSION   BG115319
VERSION     BG115319.1  GI:12608825
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 831)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10147  row: d  column: 19
            High quality sequence stop: 731.
FEATURES             Location/Qualifiers
     source          1. .831
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4416354"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_88"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             39.3%;  Score 671.8;  DB 12;  Length 831;
  Best Local Similarity   97.4%;  Pred. No. 7e-162;
  Matches  713;  Conservative    0;  Mismatches   15;  Indels    4;  Gaps    3;

Qy        820 CTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGT 879
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGT 60

Qy        880 GGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGC 939
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGC 120

Qy        940 CTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCAC 999
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCAC 180

Qy       1000 TGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGA 1059
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGA 240

Qy       1060 GAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATC 1119
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATC 300

Qy       1120 AGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGAC 1179
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGAC 360

Qy       1180 GGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGT 1239
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGT 420

Qy       1240 GAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCT 1299
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCT 480

Qy       1300 GGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAA 1359
                 |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        481 GGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGG-CCAGGCACGGAGGAA 539

Qy       1360 GTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTG 1419
                 |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        540 GTGACCAGCCCCTCCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTG 599

Qy       1420 TGCTCTCCCC-GGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTA 1478
                 |||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        600 TGCTCTCCCCGGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCCGGGCTGTA 659

Qy       1479 GATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGC 1538
                 ||||||||||||     |  ||||||||||||||||||||||||||||||| | | |
Db        660 GATCAGGTTTGCC--ACTTGATGAGAAAGATGTTCCAGTGAGGAACCCTGATCTCTAGCC 717

Qy       1539 CCCAAACACCCA 1550
                 |  ||||  |||
Db        718 CAAAAACCACCA 729


RESULT 8
BX401619/c
```

```
LOCUS       BX401619               1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX401619 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
            Homo sapiens cDNA clone CS0DL005YL19 3-PRIME, mRNA sequence.
ACCESSION   BX401619
VERSION     BX401619.1  GI:30631999
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            8942.f For more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DL005CF10NP1&cluster=8942.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DL005CF10NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DL005YL19"
                     /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
                     25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
```

```
  Query Match             39.3%;  Score 670.4;  DB 13;  Length 1201;
  Best Local Similarity   84.6%;  Pred. No. 2e-161;
  Matches  798;  Conservative   37;  Mismatches   92;  Indels   16;  Gaps    9;

Qy        747 TAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGG 806
              |: |  :||||||| |||||||||:|:||||||||||||||||||  :||||||||
Db        958 TRWGGTKTTTCCCAGGTTCATCCAGSGSAGTCCCGCCGCCTCCCAGCAGGCKTACTGGGG 899

Qy        807 CTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTT-TATC 865
              ||:|||||||||| ||||||||   || ||||||||||||| | ||||| | || |  |
Db        898 -TGSTCCCTATAGGTCCTTTAT---GGTTCCCGAGCAGGAGTGGTGGCAGTTTTTATCCC 843

Qy        866 CCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCC 925
              |||||||||||| |||||||| ||: |||||||||||||||||||:|||||||| |: ||
Db        842 CCCGCCCAGGCATGGGGCGCCTTCWCCGGCAAGAAGGGGCAGCACWTCAAACAGTTYCCC 783

Qy        926 CGGTTTGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGT 985
              ||:|| ||||||||: ||  ||||:| |||||||||||||||  ||  |||||:| | |
Db        782 CGKTTGGCCAGCGCYSCCTGCAAGWTGGCACCACCCGAAACACTGGATSCCAAAKTGCTT 723

Qy        986 ATGGTTATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGC 1045
              || |   |  ||  |||||||||||||||||:| |||||:|||||||: ||| |:|| |
Db        722 TTGTTGTNCTSCATGGGACCGCCAGAGGCCCAWTACAAGGYTCAGGGARCAATTTWTGTC 663

Qy       1046 AAACTCAAGGAGGAG-AACTTCTTTGGT--CCCAAGGAGGAAGTGAAGCTGGAGACCCAC 1102
              |||||||||:|::|:: |||||||||:|:  || |:||:|||||||||||||||||||||
Db        662 AAACTCAARGRRGRRDAACTTCTTTKGKTCCCRARGGRGGAAGTGAAGCTGGAGACCCAC 603

Qy       1103 ATACGTGTG---CCAGCATCAGCAGCTGGCCGGG--TCATTGGCAAAGGTGGAAAAACGG 1157
                  |||   |  |||  ||||||| ||| ||  |   |:: || || :||||| |||
Db        602 GATTCGGTGTTCCAGGCACTAGCAGCTTGCCNGGGCTANNTKSAAANGGGKGAAAANCGG 543

Qy       1158 TGAAC--GAGTTGCAGAATTTGACGGCA-GCTGAGGTGGTAGTACCAAGAGACCAGACCC 1214
              | |   | |||: |:| |||:  |||| ||  : |  | :||||||||||||||||||||
Db        542 GGTANCAGNGTTSAARANTTTRCAGGCAGGCNTRNGGTGGWGTACCAAGAGACCAGACCC 483

Qy       1215 CTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGG 1274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 CTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGG 423

Qy       1275 CTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGA 1334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 CTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGA 363

Qy       1335 GTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGA 1394
              |||||||||||||||||||||||||||:|||||:|||||||||||||| ||||||||||
Db        362 GTAACCAGGCCCAGGCACGGAGGAAGTGMCCAGCSCCTCCCTGTCCCTTCGAGTCCAGGA 303

Qy       1395 CAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAAT 1454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 CAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAAT 243

Qy       1455 CCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCC 1514
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 CCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCC 183

Qy       1515 AGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGC 1574
              |||||||||||||||| ||||||||||:|||||||||||||||||||||||||| |||
Db        182 AGTGAGGAACCCTGATCTCTCAGCCCCAAAMACCCACCCAATTGGCCCAACACTGTCTGC 123

Qy       1575 CCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAG 1634
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        122 CCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAG 63

Qy       1635 CTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGA 1677
              |||||||||||||||||||||||||    | |||| ||| || |
Db         62 CTCTCCAGGCCCCACCAAGAGGGTGANNCNCACCGCAGGGGTA 20
```

```
RESULT 9
BQ225582
LOCUS       BQ225582                712 bp    mRNA    linear   EST 02-MAY-2002
DEFINITION  AGENCOURT_7593363 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020837
            5', mRNA sequence.
ACCESSION   BQ225582
VERSION     BQ225582.1  GI:20406982
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13225  row: j  column: 06
            High quality sequence stop: 625.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6020837"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
```

/clone_lib="NIH_MGC_70"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."

ORIGIN

Query Match 39.2%; Score 669; DB 13; Length 712;
Best Local Similarity 99.3%; Pred. No. 3.4e-161;
Matches 669; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1031 GGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAG 1090
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAG 60

Qy       1091 CTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGA 1150
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGA 120

Qy       1151 AAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAG 1210
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAG 180

Qy       1211 ACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAG 1270
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAG 240

Qy       1271 ATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGA 1330
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGA 300

Qy       1331 CAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCC 1390
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        301 CAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCCCTTCGAGTCC 360

Qy       1391 AGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGG 1450
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGG 420

Qy       1451 GAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATG 1510
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATG 480

Qy       1511 TTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGCCCAACACTGT 1570
             |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        481 TTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACCCAATTGGCCCAACACTGT 540

Qy       1571 NTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGGATTGTTTAAA 1630
             |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        541 CTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTTAAACGTGGATTGTTTAAA 600

Qy       1631 GAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGAAAAATAAAAT 1690
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGAAAAATAAAAT 660

Qy       1691 TTCCTTCAGGTTTT 1704
             ||||||||||||||
Db        661 TTCCTTCAGGTTTT 674
```

RESULT 10
BF984962
LOCUS BF984962 896 bp mRNA linear EST 23-JAN-2001
DEFINITION 602308484F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4399556 5', mRNA sequence.
ACCESSION BF984962
VERSION BF984962.1 GI:12387774
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 896)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10103 row: h column: 21
High quality sequence stop: 652.
FEATURES Location/Qualifiers
source 1. .896
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4399556"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_88"
/note="Organ: small intestine; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
ORIGIN

Query Match 38.8%; Score 662; DB 10; Length 896;
Best Local Similarity 93.4%; Pred. No. 2.5e-159;
Matches 727; Conservative 0; Mismatches 40; Indels 11; Gaps 3;

```
Qy        583 AGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTG 642
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTG 60

Qy        643 TTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGC 702
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGC 120

Qy        703 TGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 756
              |||||||||       ||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGCCATGAGCCTGCAGTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 180

Qy        757 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 816
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 240

Qy        817 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 300

Qy        877 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 360

Qy        937 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 996
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 420

Qy        997 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1056
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 480

Qy       1057 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1116
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 540

Qy       1117 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 600
```

```
Qy       1177 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1236
             ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||| |||||
Db        601 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGAGCCCTGATGAGAACGACCACGTCAT 660

Qy       1237 CGTGAAA----ATCATCGGACATTTCTATGCCAG-TCAGATGGCTCAACGGAAGATCCGA 1291
             |||||||          | |||||||||||||| |||||||||||  | |    |
Db        661 CGTGAAACATCCATCCGGAACATTTCTATGCCAGCTCAGATGGCTCACGGAAGATCCGAG 720

Qy       1292 GACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGG 1349
              || ||||||||||||||||| |         ||| |||| |  |  || | | | ||
Db        721 AACTTCCTGGCCCAGGTTAAGCCGGCAGCATCGAAAGGACGGGTTCCCCGGCCGCCGG 778


RESULT 11
BX372273/c
LOCUS       BX372273                891 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX372273 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
            Homo sapiens cDNA clone CS0DL005YL19 3-PRIME, mRNA sequence.
ACCESSION   BX372273
VERSION     BX372273.1  GI:30434053
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 891)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 8942.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0BAI043ZG11_CS04119_1&cluster=8942.f.
            Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0BAI043ZG11_CS04119_1.
FEATURES             Location/Qualifiers
     source          1. .891
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DL005YL19"
                     /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
                     25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             38.7%;  Score 660.2;  DB 13;  Length 891;
  Best Local Similarity   94.1%;  Pred. No. 7.1e-159;
  Matches  765;  Conservative    0;  Mismatches   33;  Indels   15;  Gaps    7;

Qy        126 AGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTC--AGCCCCGCCAGGGCTCACCTGT 183
              | ||  |||  | ||||  ||| ||||||||||||    |||  |  ||| |||||||||
Db        810 AAAAGTGGCCACAAGGGTCCTTGGGCTCTCGGGGTCCAGCCCCGCCAGGGGATCACCTGT 751

Qy        184 GGCAGCGGGGGCCCCAGCCAAGC-AGCAGCAAGTGGACATCCCCCTTCGGCTCCT-GGTG 241
              |  |  |||||||||||| ||| | ||||||||||||||||||||||||||||| ||||
Db        750 GCAACGGGGGGCCCCAGGCAAACAAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGGTG 691

Qy        242 CCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAAC-ATCAC 300
              ||||| ||| |||||||| ||||||| | |||||||||||||| |||||||||| |||||
Db        690 CCCACACAGAATGTGGGTACCATTATGGCCAAGGAGGGGGCCCCCATCCGCAACAATCAC 631

Qy        301 AAAACAGACCCAGTCC-AAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAA 359
              |||||||||||||||| |||||||||||||||||||||| |||||||||||||| ||||||
Db        630 AAAACAGACCCAGTCCAAAGATAGACGTGCATAGGAAGTAGAACGCAGGTGCACCTGAAA 571

Qy        360 AAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGG 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        570 AAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGG 511

Qy        420 AGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        510 AGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCC 451

Qy        480 TGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGG 539
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        450 TGGCCCATAATACCTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGG 391

Qy        540 TAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACA 599
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        390 TAGAGCAAGATACCGAGACAAAAATCCCCATCTCCTCGTTGCAAGACCTTACCCTTTACA 331

Qy        600 ACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGG 659
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        330 ACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGG 271

Qy        660 AAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC------T 713
              |||||||||||||||||||||||||| |||||||||||||||||||||||||      |
Db        270 AAATAATGAAGAAAGTTCGGGAGGCCTGTGAGAATGATGTGGCTGCCATGAGCCTGCAGT 211

Qy        714 CTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCG 773
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 CTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCG 151

Qy        774 CAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGG 833
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||   ||
Db        150 CAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTAT---GG 94

Qy        834 CTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCG 893
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 CTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCG 34

Qy        894 GCAAGAAGGGGCAGCACATCAAACAGCTCTCCC 926
              |||||||||||||||||||||||||||||||||
Db         33 GCAAGAAGGGGCAGCACATCAAACAGCTCTCCC 1


RESULT 12
BQ651665
LOCUS       BQ651665                955 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8297761 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269842
            5', mRNA sequence.
ACCESSION   BQ651665
VERSION     BQ651665.1  GI:21775837
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 955)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2444  row: a  column: 11
            High quality sequence stop: 588.
```

```
FEATURES             Location/Qualifiers
     source          1. .955
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6269842"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match           38.3%;  Score 653.4;  DB 13;  Length 955;
  Best Local Similarity  98.4%;  Pred. No. 4.2e-157;
  Matches  660;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy         30 CCCGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 89
              ||  ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 CCAGGCAAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 64

Qy         90 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 124

Qy        150 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 184

Qy        210 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 244

Qy        270 GCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 304

Qy        330 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGG 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGG 364

Qy        390 GCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCA 424

Qy        450 AAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCA 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 AAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCA 484

Qy        510 TTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCA 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 TTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCA 544

Qy        570 TCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 TCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGG 604

Qy        630 CCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATG 689
              |||||||||||| |||||||| | || ||||||||||||||||||||||   |||||
Db        605 CCATCGAGAATTGCTGCAGGGCCCACCACGAAATAATGAAGAAAGTTCGGGAGCGCTATG 664

Qy        690 AGAATGATGTG 700
              |||||| |||
Db        665 CGAATGACGTG 675
```

```
RESULT 13
BM006191
LOCUS       BM006191                764 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  603614013F1 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5433698 5',
            mRNA sequence.
ACCESSION   BM006191
VERSION     BM006191.1  GI:16520545
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 764)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1909  row: j  column: 03
            High quality sequence stop: 756.
FEATURES             Location/Qualifiers
     source          1. .764
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5433698"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match           37.9%;  Score 647.2;  DB 12;  Length 764;
  Best Local Similarity  96.7%;  Pred. No. 1.5e-155;
  Matches  727;  Conservative    0;  Mismatches   13;  Indels   12;  Gaps    6;

Qy         30 CCCGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 89
              ||  ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          7 CCAGGCAAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 66

Qy         90 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 126

Qy        150 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 209
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db        127 GCTCTCGGGGTCAGCCCCGCCA-GGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 185

Qy        210 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 245

Qy        270 GCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 305

Qy        330 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCC-TGAG 388
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| |
```

```
Db        306 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTTGAG 365

Qy        389 GGCTGCTCCTCCGCTTGTAA-GATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACAC 447
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        366 GGCTGCTCCTCCGCTTGTAACGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACAC 425

Qy        448 CAAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCT 507
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CAAAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCT 485

Qy        508 CATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCAC 567
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 CATTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCAC 545

Qy        568 CATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGG 627
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 CATCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGG 605

Qy        628 GGCCATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTA 687
              |||||||||||||||| |||||||||||||| ||| ||||||||||||||||||||||||
Db        606 GGCCATCGAGAATTGTCGCAGGGCCGAGCAGGACATACTGAAGAAAGTTCGGGAGGCCTA 665

Qy        688 TGAGAATGATGTGGCTGCCATGAGCTCTC-------ACCTGATCCCT-GGCCTGAACCTG 739
              ||||||||||||||||  |||||||  |       || |||||||| ||| ||||||||
Db        666 TGAGAATGATGTGGCTTGCATGAGCCTGCAGTCTCAACTTGATCCCTGGGCTTGAACCTG 725

Qy        740 GCTGCTGTA-GGTCTTTTCCCAGCTTCATCCA 770
              ||||||||| |||||||||||||||||||||||
Db        726 GCTGCTGTAGGGTCTTTTCCCAGCTTCATCCA 757
```

RESULT 14
BQ643920

```
LOCUS       BQ643920                983 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8303966 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269413
            5', mRNA sequence.
ACCESSION   BQ643920
VERSION     BQ643920.1  GI:21768092
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 983)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2442  row: o  column: 14
            High quality sequence stop: 491.
FEATURES             Location/Qualifiers
     source          1. .983
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6269413"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN
```

```
  Query Match           36.7%;  Score 627.6;  DB 13;  Length 983;
  Best Local Similarity   96.3%;  Pred. No. 1.9e-150;
  Matches  653;  Conservative    0;  Mismatches   24;  Indels    1;  Gaps    1;

Qy         30 CCCGGGGAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 89
              || ||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 CCAGGCAAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCT 64

Qy         90 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CCTACATCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTG 124

Qy        150 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 GCTCTCGGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGC 184

Qy        210 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 AGCAAGTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTG 244

Qy        270 GCAAGGAGGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCAAGGAGGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGC 304

Qy        330 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGG 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ATAGGAAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGG 364

Qy        390 GCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GCTGCTCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCA 424

Qy        450 AAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCA 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 AAACGGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCA 484

Qy        510 TTGGCAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCA 569
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        485 TTGGCAAGGAAGGACGGAACCTGAAGAAAGTAGAGCAAGATACCGAGACAAAAATCACCA 544

Qy        570 TCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGG 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 TCTCCTCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGG 604

Qy        630 CC-ATCGAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTAT 688
              || |||||||||||||||||| |||||||||||||  ||||||| || ||  | | | |
Db        605 CCAATCGAGAATTGTTGCAGGCCCGAGCAGGAAATATGGAAGAAAGGTCCGGGAGGCCCT 664

Qy        689 GAGAATGATGTGGCTGCC 706
              | | |||  |  | ||
Db        665 AATGAAAATGATGGGGGC 682
```

RESULT 15
AK088465

```
LOCUS       AK088465                4100 bp    mRNA    linear   HTC 20-SEP-2003
DEFINITION  Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
            full-length enriched library, clone:E430018F06 product:insulin-like
            growth factor 2, binding protein 3, full insert sequence.
ACCESSION   AK088465
VERSION     AK088465.1  GI:26353479
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
```

```
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 4100)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .4100
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="FANTOM_DB:E430018F06"
                     /db_xref="MGI:2427362"
                     /db_xref="taxon:10090"
                     /clone="E430018F06"
                     /cell_type="thymic cells"
                     /tissue_type="thymus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="2 days neonate"
     CDS             243. .1982
                     /note="unnamed protein product; insulin-like growth factor
                     2, binding protein 3 (MGD|MGI:1890359, GB|NM_023670,
                     evidence: BLASTN, 99%, match=1740)
                     putative"
                     /codon_start=1
                     /protein_id="BAC40370.1"
                     /db_xref="GI:26353480"
                     /translation="MNKLYIGNLSDHAGPADLESVFKDAKIPVAGPFLVKTGYAFVDC
                     PDEGWALKAIEALSGKMELHGKPMEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLV
                     QYGVVESCEQVNTDSETAVVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAA
                     QQNPSPQLRGRRGPGQRGSSRQASPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATI
                     RNITKQTQSKIDVHRKENTGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEE
                     IPLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGSVE
                     TCAKAEEEIMKKIRESYENDIASMNLQANLIPGLNLNALGLFPPTSGMPPPTSGPPST
                     LTPPYPQFEQSETETVHLFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAK
                     VRMVIITGPPEAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGK
                     TVNELQSLSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQRKIQEILTQVKQHQQQ
                     KALQSGPPQSRRK"
     polyA_signal    4079. .4084
                     /note="putative"
     polyA_site      4100
                     /note="putative"
ORIGIN

  Query Match           36.7%;  Score 627;  DB 11;  Length 4100;
  Best Local Similarity  68.3%;  Pred. No. 6e-150;
  Matches  886;  Conservative    0;  Mismatches  405;  Indels    6;  Gaps    1;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
Db        644 AGCCTTAGACAAACTGAATGGATTCCAGTTAGAGAACTTCACCTTGAAGGTTGCCTACAT 703

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
Db        704 CCCAGATGAAACTGCTGCCCAGCAAAATCCCTCGCCACAGCTCCGGGGGCGCCGGGGGCC 763

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
Db        764 AGGGCAGCGGGGCTCATCCAGGCAGGCGTCTCCAGGATCGGTGTCCAAGCAGAAACCCTG 823

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
Db        824 TGACCTGCCACTGCGCCTGCTGGTTCCCACCCAGTTTGTTGGAGCCATTATAGGAAAAGA 883

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
Db        884 AGGTGCTACCATTCGCAACATCACAAAGCAGACCCAGTCTAAAATCGATGTCCATCGTAA 943

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
Db        944 GGAGAATACAGGGGCCGCGGAGAAGTCCATTACTATCCTCTCTACCCCTGAAGGCACCTC 1003

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
Db       1004 TGCAGCTTGTAAGTCTATTCTGGAGATTATGCATAAGGAAGCTCAAGATATAAAATTCAC 1063

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
Db       1064 AGAGGAGATTCCCTTGAAGATCTTAGCTCACAATAACTTTGTAGGCCGTCTCATTGGTAA 1123

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
Db       1124 AGAAGGAAGAAACCTTAAAAAAATCGAGCAAGACACGGACACTAAAATCACAATATCTCC 1183
```

```
Qy       577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
             ||||| || || || || ||||| || ||  | ||||| || ||||| ||      | ||
Db      1184 ATTGCAGGAACTGACGCTGTACAATCCGGAACGCACCATTACAGTGAAAGGCAGTGTTGA 1243

Qy       637 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 696
             ||  |||  || ||| ||| ||||||||||||||||  |  ||||| | ||||| |||||
Db      1244 GACGTGTGCCAAGGCGGAGGAGGAAATAATGAAGAAGATCAGGGAGTCTTATGAAAATGA 1303

Qy       697 TGTGGCTGCCATGAGCTCTC------ACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGG 750
             | | ||| ||||||    ||      |  | ||||||||  | || |||  |||  | ||
Db      1304 TATTGCTTCCATGAATCTTCAAGCAAATTTAATCCCTGGATTAAATCTGAATGCCTTGGG 1363

Qy       751 TCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGC 810
             ||| |||||| |  | ||  |        || ||  |  |  |  |  | |         |
Db      1364 TCTGTTCCCACCCACGTCAGGGATGCCACCTCCCACCTCAGGGCCCCCTTCAACCCTGAC 1423

Qy       811 TCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGC 870
             |||       | |     |  ||    | |||  |||||  |||||  ||||||| |||||
Db      1424 TCCTCCCTACCCACAATTTGAGCAATCAGAGACGGAGACTGTGCATCTGTTTATTCCCGC 1483

Qy       871 CCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTT 930
             || | | || |||||||||| |||||| |||| || |||||||||||||| || || ||
Db      1484 CCTGTCCGTTGGCGCCATCATTGGCAAGCAGGGCCAACACATCAAACAGCTTTCTCGCTT 1543

Qy       931 TGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGT 990
             |||  | || || || ||||| || ||| ||||| |||| ||  | ||||| || |||||
Db      1544 TGCGGGAGCTTCGATTAAGATCGCTCCAGCCGAAGCACCAGATGCTAAAGTGCGGATGGT 1603

Qy       991 TATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACT 1050
             || |||||||||| |||||||| || ||||||||||||||||||| ||||| ||| |
Db      1604 GATTATCACTGGACCACCAGAGGCTCAGTTCAAGGCTCAGGGAAGAATTTATGGAAAAAT 1663

Qy      1051 CAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGT 1110
             || || || |||||  || |||| || || || |||||  || ||  | |||||  | ||
Db      1664 TAAAGAAGAAAACTTTGTTAGTCCTAAAGAAGAGGTGAAACTTGAAGCTCACATCAGAGT 1723

Qy      1111 GCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCA 1170
             |||  | |  || |||||  | || |||||  ||||| || |||||||||| ||| | ||
Db      1724 GCCGTCCTTTGCTGCTGGCAGAGTTATTGGGAAAGGAGGCAAAACGGTGAATGAGCTCCA 1783

Qy      1171 GAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCA 1230
             || |||  |    ||||| || || || ||  | |||||||  |||||||||| || ||
Db      1784 GAGTTTATCAAGTGCTGAAGTTGTCGTCCCCCGTGACCAGACACCTGATGAGAATGATCA 1843

Qy      1231 GGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCG 1290
             ||  | || ||||| |  || || ||||||||  | ||| | || ||  |||| || |
Db      1844 AGTAGTTGTCAAAATAACTGGCCACTTCTATGCTTGCCAGGTTGCCCAGAGGAAAATTCA 1903

Qy      1291 AGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAG 1327
             || || ||| | ||||| |||||||| || ||| ||
Db      1904 GGAAATTCTGACTCAGGTAAAGCAGCACCAGCAGCAG 1940
```

Search completed: July 13, 2004, 21:40:34
Job time : 4603 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 14, 2004, 01:45:09 ; Search time 1662 Seconds
(without alignments)
5008.780 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 1708
Sequence: 1 agggacgctgccgcaccgcc.....atttccttcaggttttaaaa 1708

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3183909 seqs, 2436941669 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6367818

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1708	100.0	1708	9	US-09-899-651-5	Sequence 5, Appli
	2	1672	97.9	1946	9	US-09-899-651-7	Sequence 7, Appli
	3	741	43.4	1182	9	US-09-764-864-327	Sequence 327, App
	4	588	34.4	1129	9	US-09-764-864-745	Sequence 745, App
c	5	365	21.4	1738	16	US-10-062-674-2082	Sequence 2082, Ap
	6	125	7.3	831	13	US-10-027-632-157364	Sequence 157364,
	7	125	7.3	831	16	US-10-027-632-157364	Sequence 157364,
	8	101	5.9	762	16	US-10-027-632-20264	Sequence 20264, A
	9	60	3.5	60	10	US-09-908-975-15619	Sequence 15619, A
	10	60	3.5	60	10	US-09-908-975-31745	Sequence 31745, A
	11	60	3.5	60	10	US-09-908-975-31873	Sequence 31873, A
	12	50	2.9	50	16	US-10-131-827-2939	Sequence 2939, Ap
	13	31	1.8	2224	9	US-09-873-637-1	Sequence 1, Appli
	14	29	1.7	412	16	US-10-062-674-1244	Sequence 1244, Ap
c	15	24	1.4	24	9	US-09-873-637-38	Sequence 38, Appl
	16	23	1.3	23	9	US-09-873-637-36	Sequence 36, Appl
	17	23	1.3	65	10	US-09-908-975-30184	Sequence 30184, A
c	18	22	1.3	495	13	US-10-085-783A-16926	Sequence 16926, A
c	19	22	1.3	495	16	US-10-242-535A-16926	Sequence 16926, A
	20	22	1.3	901	13	US-10-425-114-9333	Sequence 9333, Ap
	21	22	1.3	973	13	US-10-424-599-15683	Sequence 15683, A
	22	22	1.3	1113	13	US-10-424-599-141668	Sequence 141668,
c	23	21	1.2	378	13	US-10-424-599-97191	Sequence 97191, A
c	24	21	1.2	1954	13	US-10-425-114-11930	Sequence 11930, A
	25	21	1.2	1995	13	US-10-425-114-10752	Sequence 10752, A
c	26	21	1.2	2073	13	US-10-424-599-47612	Sequence 47612, A
	27	21	1.2	2289	13	US-10-424-599-46916	Sequence 46916, A
	28	20	1.2	238	15	US-10-029-386-16656	Sequence 16656, A
c	29	20	1.2	520	13	US-10-027-632-270463	Sequence 270463,
c	30	20	1.2	520	16	US-10-027-632-270463	Sequence 270463,
	31	20	1.2	599	15	US-10-029-386-2956	Sequence 2956, Ap
c	32	20	1.2	626	13	US-10-027-632-115479	Sequence 115479,
c	33	20	1.2	626	16	US-10-027-632-115479	Sequence 115479,
c	34	20	1.2	1465	9	US-09-925-302-70	Sequence 70, Appl
c	35	20	1.2	1465	13	US-09-925-302-70	Sequence 70, Appl
c	36	20	1.2	1505	17	US-10-437-963-6811	Sequence 6811, Ap
	37	20	1.2	1740	9	US-09-735-705-347	Sequence 347, App
	38	20	1.2	1740	9	US-09-850-716A-347	Sequence 347, App
	39	20	1.2	1740	9	US-09-897-778-347	Sequence 347, App
	40	20	1.2	1740	13	US-10-007-700-347	Sequence 347, App
	41	20	1.2	1740	15	US-10-117-982-347	Sequence 347, App
	42	20	1.2	1740	15	US-10-117-982-478	Sequence 478, App
	43	20	1.2	1740	15	US-10-117-982-483	Sequence 483, App
	44	20	1.2	1740	16	US-10-313-986-347	Sequence 347, App
	45	20	1.2	1740	16	US-10-313-986-478	Sequence 478, App

ALIGNMENTS

RESULT 1
US-09-899-651-5
; Sequence 5, Application US/09899651
; Patent No. US20020111470A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
; TITLE OF INVENTION: Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/899,651
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/061,709
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 5
; LENGTH: 1708
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-899-651-5

Query Match 100.0%; Score 1708; DB 9; Length 1708;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGGGACGCTGCCGCACCGCCCCAGTTTACCCCGGGGAGCCATCATGAAGCTGAATGGCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGGACGCTGCCGCACCGCCCCAGTTTACCCCGGGGAGCCATCATGAAGCTGAATGGCCA 60

Qy         61 CCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCACAGGG 120
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCACAGGG 120

Qy        121 ACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACC 180

Qy        181 TGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGT 240

Qy        241 GCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCAC 300

Qy        301 AAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAA 360

Qy        361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGA 420

Qy        421 GATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCT 480

Qy        481 GGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGT 540

Qy        541 AGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAA 600

Qy        601 CCCTGAGAGGACCATCACTGTGAAGGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCCTGAGAGGACCATCACTGTGAAGGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGA 660

Qy        661 AATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCTCTCACCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCTCTCACCT 720

Qy        721 GATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCC 780

Qy        781 GCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGA 840

Qy        841 GCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAA 900

Qy        901 GGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACC 960

Qy        961 CGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATT 1020

Qy       1021 CAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080

Qy       1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGG 1140

Qy       1141 CAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACC 1200

Qy       1201 AAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTA 1260

Qy       1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCA 1320

Qy       1321 TCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCC 1380

Qy       1381 CTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGA 1440

Qy       1441 GAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATT 1500

Qy       1501 GAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGC 1560

Qy       1561 CCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGG 1620

Qy       1621 ATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA 1680

Qy       1681 AAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||
Db       1681 AAAATAAAATTTCCTTCAGGTTTTAAAA 1708
```

```
RESULT 2
US-09-899-651-7
; Sequence 7, Application US/09899651
; Patent No. US20020111470A1
; GENERAL INFORMATION:
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT:  Gure, Ali
;  APPLICANT:  Tsang, Solam
;  APPLICANT:  Stockert, Elisabeth
;  APPLICANT:  Jager, Elke
;  APPLICANT:  Knuth, Alexander
;  APPLICANT:  Old, Lloyd J.
;  TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer
;  TITLE OF INVENTION:  Associated
;  TITLE OF INVENTION:  Antigen, The Antigens Per Se, And Uses Thereof
;  FILE REFERENCE: LUD 5538
;  CURRENT APPLICATION NUMBER: US/09/899,651
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/09/061,709
;  PRIOR FILING DATE: 1998-04-17
;  NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 7
;   LENGTH: 1946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-899-651-7

  Query Match             97.9%;  Score 1672;  DB 9;  Length 1946;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1672;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        275 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 334
Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
Db        335 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 394
Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
Db        395 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 454
Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
Db        455 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 514
Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
Db        515 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 574
Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
Db        575 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 634
Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
Db        635 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 694
Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
Db        695 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 754
Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
Db        755 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 814
Qy        577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
Db        815 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 874
Qy        637 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 696
Db        875 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 934
Qy        697 TGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 756
Db        935 TGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 994
Qy        757 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 816
Db        995 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 1054
Qy        817 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 876
Db       1055 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 1114
Qy        877 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 936
Db       1115 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 1174
Qy        937 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 996
Db       1175 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 1234
Qy        997 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1056
Db       1235 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1294
Qy       1057 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1116
Db       1295 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1354
Qy       1117 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1176
Db       1355 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1414
```

```
Qy       1177 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1236
Db       1415 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1474
Qy       1237 CGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACAT 1296
Db       1475 CGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACAT 1534
Qy       1297 CCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAG 1356
Db       1535 CCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAG 1594
Qy       1357 GAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGA 1416
Db       1595 GAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGA 1654
Qy       1417 GTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTG 1476
Db       1655 GTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTG 1714
Qy       1477 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCA 1536
Db       1715 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCA 1774
Qy       1537 GCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNT 1596
Db       1775 GCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNT 1834
Qy       1597 AGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGG 1656
Db       1835 AGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGG 1894
Qy       1657 GTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
Db       1895 GTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1946
```

RESULT 3
US-09-764-864-327
; Sequence 327, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 327
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-327

Query Match 43.4%; Score 741; DB 9; Length 1182;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 991; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
Db        136 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 195
Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
Db        196 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 255
Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
Db        256 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 315
```

```
Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 375

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 435

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 495

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 555

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 615

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 675

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 735

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 795

Qy       1373 CCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1432
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 CCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 855

Qy       1433 AGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCC 1492
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        856 AGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCC 915

Qy       1493 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACC 1552
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        916 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACC 975

Qy       1553 CAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTT 1612
              |||||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db        976 CAATTGGCCCAACACTGTCTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTT 1035

Qy       1613 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1095

Qy       1673 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||||||||||
Db       1096 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1131


RESULT 4
US-09-764-864-745
; Sequence 745, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 745
;   LENGTH: 1129
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-864-745

  Query Match             34.4%;  Score 588;  DB 9;  Length 1129;
  Best Local Similarity   99.2%;  Pred. No. 3.9e-297;
  Matches  988;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        103 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 162

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 222

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 282

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        283 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 342

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
              |||||||||||||||||||||||||||||||||||||||| |||||| |||||||| ||
Db        343 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGKTATCATYACTGGACCGCCARAG 402

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 462

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 522

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 582

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 642

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 702

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 762

Qy       1373 CCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1432
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 822

Qy       1433 AGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCC 1492
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        823 AGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCC 882

Qy       1493 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACC 1552
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        883 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACC 942

Qy       1553 CAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTT 1612
              |||||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db        943 CAATTGGCCCAACACTGTCTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTT 1002

Qy       1613 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1062

Qy       1673 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||||||||||
```

```
Db       1063 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1098


RESULT 5
US-10-062-674-2082/c
; Sequence 2082, Application US/10062674
; Publication No. US20040005559A1
; GENERAL INFORMATION:
;  APPLICANT: Loring, Jeanne F.; Kaser, Matthew R.
;  TITLE OF INVENTION: MARKERS OF NEURONAL DIFFERENTIATION AND MORPHOGENESIS
;  FILE REFERENCE: PA-0026-1 CIP
;  CURRENT APPLICATION NUMBER: US/10/062,674
;  CURRENT FILING DATE:  2002-01-30
;  PRIOR APPLICATION NUMBER: US 09/625,102
;  PRIOR FILING DATE: 2000-07-24
;  NUMBER OF SEQ ID NOS: 2217
;  SOFTWARE: PERL Program
; SEQ ID NO 2082
;   LENGTH: 1738
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20040005559A1 457674.1
US-10-062-674-2082

  Query Match             21.4%;  Score 365;  DB 16;  Length 1738;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-180;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1728 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 1669

Qy          97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1668 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 1609

Qy         157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1608 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 1549

Qy         217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1548 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 1489

Qy         277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1488 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 1429

Qy         337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1428 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 1369

Qy         397 CTCCG 401
               |||||
Db        1368 CTCCG 1364


RESULT 6
US-10-027-632-157364
; Sequence 157364, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 157364
;   LENGTH: 831
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-157364

  Query Match             7.3%;  Score 125;  DB 13;  Length 831;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-54;
  Matches  125;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         454 GGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGG 513
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         344 GGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGG 403

Qy         514 CAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTC 573
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         404 CAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTC 463

Qy         574 CTCGT 578
               |||||
Db         464 CTCGT 468


RESULT 7
US-10-027-632-157364
; Sequence 157364, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 157364
;   LENGTH: 831
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-157364

  Query Match             7.3%;  Score 125;  DB 16;  Length 831;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-54;
  Matches  125;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        454 GGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGG 513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 GGCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGG 403

Qy        514 CAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTC 573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        404 CAAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTC 463

Qy        574 CTCGT 578
              |||||
Db        464 CTCGT 468
```

RESULT 8
US-10-027-632-20264
; Sequence 20264, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20264
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-20264

Query Match 5.9%; Score 101; DB 16; Length 762;
Best Local Similarity 100.0%; Pred. No. 6.2e-42;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        354 CTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGA 413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGA 60

Qy        414 TCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACG 454
              |||||||||||||||||||||||||||||||||||||||||
Db         61 TCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACG 101
```

RESULT 9
US-09-908-975-15619
; Sequence 15619, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15619
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-15619

Query Match 3.5%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1243 AATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGC 1302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGC 60
```

RESULT 10
US-09-908-975-31745
; Sequence 31745, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31745
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-31745

Query Match 3.5%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1470 CGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGA 1529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGA 60
```

RESULT 11
US-09-908-975-31873
; Sequence 31873, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20

```
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31873
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-908-975-31873

  Query Match             3.5%;  Score 60;  DB 10;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-20;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1470 CGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGA 1529
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGA 60


RESULT 12
US-10-131-827-2939
; Sequence 2939, Application US/10131827
; Publication No. US20040009479A1
; GENERAL INFORMATION:
;  APPLICANT: Wohlgemuth, Jay
;  APPLICANT:  Fry, Kirk
;  APPLICANT:  Woodward, Robert
;  APPLICANT:  Ly, Ngoc
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
;  TITLE OF INVENTION:  CHRONIC INFLAMMATORY DISEASES
;  FILE REFERENCE: 506612000120
;  CURRENT APPLICATION NUMBER: US/10/131,827
;  CURRENT FILING DATE:  2002-09-06
;  PRIOR APPLICATION NUMBER: US 10/006,290
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: US 60/296,764
;  PRIOR FILING DATE: 2001-06-08
;  NUMBER OF SEQ ID NOS: 9090
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2939
;   LENGTH: 50
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-131-827-2939

  Query Match             2.9%;  Score 50;  DB 16;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-15;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1652 AGAGGGTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGT 1701
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGAGGGTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGT 50


RESULT 13
US-09-873-637-1
; Sequence 1, Application US/09873637
; Patent No. US20020061543A1
; GENERAL INFORMATION:
;  APPLICANT: Ross, Jeffrey
;  TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
;  TITLE OF INVENTION:  (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
;  FILE REFERENCE: 960296.95131
;  CURRENT APPLICATION NUMBER: US/09/873,637
;  CURRENT FILING DATE:  2001-06-04
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2224
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-873-637-1

  Query Match             1.8%;  Score 31;  DB 9;  Length 2224;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-05;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         66 TGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              |||||||||||||||||||||||||||||||
Db        561 TGGAGAACCATGCCCTGAAGGTCTCCTACAT 591


RESULT 14
US-10-062-674-1244
; Sequence 1244, Application US/10062674
; Publication No. US20040005559A1
; GENERAL INFORMATION:
;  APPLICANT: Loring, Jeanne F.; Kaser, Matthew R.
;  TITLE OF INVENTION: MARKERS OF NEURONAL DIFFERENTIATION AND MORPHOGENESIS
;  FILE REFERENCE: PA-0026-1 CIP
;  CURRENT APPLICATION NUMBER: US/10/062,674
;  CURRENT FILING DATE:  2002-01-30
;  PRIOR APPLICATION NUMBER: US 09/625,102
;  PRIOR FILING DATE: 2000-07-24
;  NUMBER OF SEQ ID NOS: 2217
;  SOFTWARE: PERL Program
; SEQ ID NO 1244
;   LENGTH: 412
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: GenBank ID No. US20040005559A1 g1595304
US-10-062-674-1244

  Query Match             1.7%;  Score 29;  DB 16;  Length 412;
  Best Local Similarity   100.0%;  Pred. No. 0.00033;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        599 AACCCTGAGAGGACCATCACTGTGAAGGG 627
              |||||||||||||||||||||||||||||
Db        308 AACCCTGAGAGGACCATCACTGTGAAGGG 336


RESULT 15
US-09-873-637-38/c
; Sequence 38, Application US/09873637
; Patent No. US20020061543A1
; GENERAL INFORMATION:
;  APPLICANT: Ross, Jeffrey
;  TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
;  TITLE OF INVENTION:  (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
;  FILE REFERENCE: 960296.95131
;  CURRENT APPLICATION NUMBER: US/09/873,637
;  CURRENT FILING DATE:  2001-06-04
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
;   LENGTH: 24
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Oligonucleotide primer
US-09-873-637-38

  Query Match             1.4%;  Score 24;  DB 9;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.14;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        508 CATTGGCAAGGAAGGACGGAACCT 531
              ||||||||||||||||||||||||
Db         24 CATTGGCAAGGAAGGACGGAACCT 1
```

Search completed: July 14, 2004, 05:27:47
Job time : 1666 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 21:40:40 ; Search time 700 Seconds
(without alignments)
10365.603 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 1708
Sequence: 1 agggacgctgccgcaccgcc.....atttccttcaggttttaaaa 1708

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1708	100.0	1708	3	AAZ36151	Aaz36151 DNA encod
	2	1672	97.9	1946	3	AAZ36153	Aaz36153 An altern
	3	741	43.4	1182	4	AAS26148	Aas26148 Human cDN
	4	741	43.4	1182	7	ABX73489	Abx73489 Human nov
	5	741	43.4	2780	4	AAK94782	Aak94782 Human ful
	6	588	34.4	1129	4	AAS26566	Aas26566 Human cDN
	7	588	34.4	1129	7	ABX73907	Abx73907 Human nov
	8	414	24.2	833	4	AAK91969	Aak91969 Human cDN
	9	414	24.2	833	4	AAK93655	Aak93655 Human cDN
	10	60	3.5	60	6	ABN42871	Abn42871 Human spl
	11	60	3.5	60	6	ABN59125	Abn59125 Human spl
	12	60	3.5	60	6	ABN58997	Abn58997 Human spl
	13	50	2.9	50	6	ABZ02948	Abz02948 Human leu
	14	31	1.8	2224	2	AAZ10617	cDNA enco
c	15	24	1.4	24	2	AAZ10625	Aaz10625 PCR prime
	16	23	1.3	23	2	AAZ10623	Aaz10623 PCR prime
c	17	23	1.3	23	3	AAZ36158	Aaz36158 PCR prime
	18	23	1.3	65	6	ABN57436	Abn57436 Mouse spl
	19	22	1.3	329	6	ABT10006	Abt10006 Human bre
	20	22	1.3	465	1	AAN50077	Aan50077 Soybean h
	21	20	1.2	456	5	AAS86149	Aas86149 DNA encod
c	22	20	1.2	526	3	AAC00956	Aac00956 Human sec
c	23	20	1.2	899	4	AAH05615	Aah05615 Human cDN
c	24	20	1.2	1465	3	AAF18051	Aaf18051 Lung canc
c	25	20	1.2	1553	9	ADC31862	Adc31862 Human nov
c	26	20	1.2	1553	9	ADE09844	Ade09844 Novel DNA
	27	20	1.2	1740	3	AAC66035	Aac66035 Human lun
	28	20	1.2	1740	6	ABL49254	Abl49254 Human lun
	29	20	1.2	1740	6	ABQ92440	Abq92440 Human lun
	30	20	1.2	1740	8	ADA28437	Ada28437 Human lun
	31	20	1.2	1743	6	ABL49299	Abl49299 Human lun
	32	20	1.2	1743	6	ABL49297	Abl49297 Human lun
	33	20	1.2	1743	6	ABQ92483	Abq92483 Human lun
	34	20	1.2	1743	6	ABQ92485	Abq92485 Human lun
	35	20	1.2	1743	8	ADA28540	Ada28540 Recombina
	36	20	1.2	1743	8	ADA28537	Ada28537 Recombina
	37	20	1.2	1764	6	ABL49283	Abl49283 Human lun
	38	20	1.2	1764	6	ABQ92469	Abq92469 Human lun
	39	20	1.2	1764	8	ADA28518	Ada28518 Recombina
	40	20	1.2	1985	5	AAS76779	Aas76779 DNA encod
c	41	20	1.2	2963	4	AAH17630	Aah17630 Human cDN
c	42	20	1.2	3055	6	ABX97087	Abx97087 Human NOV
	43	20	1.2	4108	5	ABV24751	Abv24751 Human pro
	44	20	1.2	4159	3	AAZ36150	Aaz36150 DNA encod
	45	20	1.2	4181	3	AAC65900	Aac65900 Human lun

ALIGNMENTS

RESULT 1
AAZ36151
ID AAZ36151 standard; DNA; 1708 BP.
XX
AC AAZ36151;
XX
DT 11-FEB-2000 (first entry)
XX
DE DNA encoding cancer associated antigen KOC-2.
XX
KW Cancer associated antigen; KOC-2; cancer; vaccine; CT7; ss.
XX
OS Homo sapiens.
XX
PN WO9954738-A1.
XX
PD 28-OCT-1999.
XX
PF 16-MAR-1999; 99WO-US005766.
XX
PR 17-APR-1998; 98US-00061709.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
XX
DR WPI; 2000-013284/01.
XX
PT Nucleotides representing cancer-associated genes, used to develop
PT products for the diagnosis, monitoring and treatment of cancers.
XX
PS Claim 55; Page 40; 44pp; English.
XX
CC The present sequence represents a cancer associated antigen gene
CC designated KOC-2. The specification also describes a cancer associated
CC antigen designated CT7. The CT7 polynucleotide was isolated from SK-MEL-
CC 37 melanoma cells. The polypeptide has some homology with MAGE-10,
CC limited to about 210 carboxy terminal amino acids. The amino terminal of
CC the protein has a repetitive pattern, with repeats rich in serine,
CC proline, glutamine and leucine, and an almost invariable core of the
CC peptide given in AAY43877. The CT7 polypeptide can be processed to
CC peptides which provoke lysis by cytolytic T cells. The polynucleotides
CC and polypeptides can be used for treating a cancerous condition and
CC screening for or diagnosing cancerous conditions. The cancer associated
CC antigens can be used as an immunogenic or vaccine composition with an
CC adjuvant, e.g. a cytokine, a saponin, or granulocyte macrophage-colony

```
CC   stimulating factor (GM-CSF)
XX
SQ   Sequence 1708 BP; 447 A; 469 C; 473 G; 314 T; 0 U; 5 Other;

  Query Match           100.0%;  Score 1708;  DB 3;  Length 1708;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1708;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGGGACGCTGCCGCACCGCCCCAGTTTACCCCGGGGAGCCATCATGAAGCTGAATGGCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGGACGCTGCCGCACCGCCCCAGTTTACCCCGGGGAGCCATCATGAAGCTGAATGGCCA 60

Qy         61 CCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCACAGGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCACAGGG 120

Qy        121 ACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACC 180

Qy        181 TGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGT 240

Qy        241 GCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCAC 300

Qy        301 AAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAA 360

Qy        361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGA 420

Qy        421 GATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCT 480

Qy        481 GGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGT 540

Qy        541 AGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAA 600

Qy        601 CCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGA 660

Qy        661 AATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCTCTCACCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCTCTCACCT 720

Qy        721 GATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCC 780

Qy        781 GCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGA 840

Qy        841 GCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAA 900

Qy        901 GGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACC 960

Qy        961 CGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATT 1020

Qy       1021 CAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080

Qy       1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGG 1140

Qy       1141 CAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACC 1200

Qy       1201 AAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTA 1260

Qy       1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCA 1320

Qy       1321 TCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCC 1380

Qy       1381 CTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGA 1440

Qy       1441 GAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATT 1500

Qy       1501 GAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGC 1560

Qy       1561 CCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGG 1620

Qy       1621 ATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA 1680

Qy       1681 AAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||
Db       1681 AAAATAAAATTTCCTTCAGGTTTTAAAA 1708


RESULT 2
AAZ36153
ID   AAZ36153 standard; DNA; 1946 BP.
XX
AC   AAZ36153;
XX
DT   11-FEB-2000  (first entry)
XX
DE   An alternative form of DNA encoding cancer associated antigen KOC-2.
XX
KW   Cancer associated antigen; KOC-2; cancer; vaccine; CT7; ss.
XX
OS   Homo sapiens.
XX
PN   WO9954738-A1.
XX
PD   28-OCT-1999.
XX
PF   16-MAR-1999;   99WO-US005766.
XX
PR   17-APR-1998;   98US-00061709.
```

XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gure A, Tsang S, Stockert E, Jager E, Knuth A, Old LJ;
XX
DR WPI; 2000-013284/01.
XX
PT Nucleotides representing cancer-associated genes, used to develop
PT products for the diagnosis, monitoring and treatment of cancers.
XX
PS Claim 55; Page 42; 44pp; English.
XX
CC The present sequence represents an alternative form of a cancer
CC associated antigen gene designated KOC-2. The specification also
CC describes a cancer associated antigen designated CT7. The CT7
CC polynucleotide was isolated from SK-MEL-37 melanoma cells. The
CC polypeptide has some homology with MAGE-10, limited to about 210 carboxy
CC terminal amino acids. The amino terminal of the protein has a repetitive
CC pattern, with repeats rich in serine, proline, glutamine and leucine, and
CC an almost invariable core of the peptide given in AAY43877. The CT7
CC polypeptide can be processed to peptides which provoke lysis by cytolytic
CC T cells. The polynucleotides and polypeptides can be used for treating a
CC cancerous condition and screening for or diagnosing cancerous conditions.
CC The cancer associated antigens can be used as an immunogenic or vaccine
CC composition with an adjuvant, e.g. a cytokine, a saponin, or granulocyte
CC macrophage-colony stimulating factor (GM-CSF)
XX
SQ Sequence 1946 BP; 502 A; 528 C; 553 G; 358 T; 0 U; 5 Other;

Query Match 97.9%; Score 1672; DB 3; Length 1946;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       275 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 334

Qy        97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       335 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 394

Qy       157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       395 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 454

Qy       217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       455 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 514

Qy       277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       515 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 574

Qy       337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       575 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 634

Qy       397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       635 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 694

Qy       457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       695 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 754

Qy       517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       755 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 814

Qy       577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       815 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 874

Qy       637 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 696
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       875 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 934

Qy       697 TGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 756
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       935 TGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 994

Qy       757 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 816
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       995 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 1054

Qy       817 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 876
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1055 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 1114

Qy       877 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 936
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1115 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 1174

Qy       937 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 996
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1175 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 1234

Qy       997 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1056
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1235 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1294

Qy      1057 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1116
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1295 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1354

Qy      1117 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1176
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1355 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1414

Qy      1177 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1236
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1415 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1474

Qy      1237 CGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACAT 1296
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1475 CGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACAT 1534

Qy      1297 CCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAG 1356
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1535 CCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAG 1594

Qy      1357 GAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGA 1416
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1595 GAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGA 1654

Qy      1417 GTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTG 1476
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1655 GTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTG 1714

Qy      1477 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCA 1536
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1715 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCA 1774

Qy      1537 GCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNT 1596
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1775 GCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNT 1834

Qy      1597 AGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGG 1656
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1835 AGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGG 1894

Qy      1657 GTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1895 GTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1946
```

RESULT 3
AAS26148
ID AAS26148 standard; cDNA; 1182 BP.
XX
AC AAS26148;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 327.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.

```
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   P-PSDB; AAU16161.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; SEQ ID NO 327; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence encodes a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

  Query Match             43.4%;  Score 741;  DB 4;  Length 1182;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches  991;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 195

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 255

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 315

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 375

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 435

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 495

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 555

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 615

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 675

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 735

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 795

Qy       1373 CCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1432
              |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db        796 CCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 855

Qy       1433 AGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCC 1492
              ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        856 AGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCC 915

Qy       1493 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACC 1552
              ||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db        916 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACC 975

Qy       1553 CAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTT 1612
              ||||||||||||||||||  ||||||||||||||||||||||  ||||||||||||||||
Db        976 CAATTGGCCCAACACTGTCTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTT 1035

Qy       1613 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1095

Qy       1673 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||||||||||
Db       1096 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1131


RESULT 4
ABX73489
ID   ABX73489 standard; DNA; 1182 BP.
XX
AC   ABX73489;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polynucleotide #317.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
```

KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-00764864.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2003-147444/14.
DR P-PSDB; ABU55229.
XX
PT New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.
XX
PS Claim 1; SEQ ID NO 327; 402pp; English.
XX
CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC human novel polynucleotides of the invention
XX
SQ Sequence 1182 BP; 343 A; 310 C; 314 G; 215 T; 0 U; 0 Other;

Query Match 43.4%; Score 741; DB 7; Length 1182;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 991; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 195

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 255

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 315

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 375

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 435

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 495

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 555

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 615

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 675

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 735

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 795

Qy       1373 CCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1432
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 CCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 855
```

```
Qy       1433 AGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCC 1492
             ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        856 AGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCC 915

Qy       1493 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACC 1552
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        916 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACC 975

Qy       1553 CAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTT 1612
             |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db        976 CAATTGGCCCAACACTGTCTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTT 1035

Qy       1613 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1095

Qy       1673 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
             |||||||||||||||||||||||||||||||||||
Db       1096 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1131
```

RESULT 5
AAK94782
ID AAK94782 standard; cDNA; 2780 BP.
XX
AC AAK94782;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 3886.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
DR P-PSDB; AAM93826.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
PS Claim 8; SEQ ID NO 3886; 1380pp + Sequence Listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesising the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a full length human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in CD-ROM format directly
CC from EPO
XX
SQ Sequence 2780 BP; 768 A; 681 C; 671 G; 660 T; 0 U; 0 Other;

```
Query Match             43.4%;  Score 741;  DB 4;  Length 2780;
Best Local Similarity   99.5%;  Pred. No. 0;
Matches  991;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 480

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 540

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 600

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 660

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 720

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 780

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 840

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 900

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 960

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1020

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1080

Qy       1373 CCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1432
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1140

Qy       1433 AGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCC 1492
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       1141 AGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCC 1200

Qy       1493 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACC 1552
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1201 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACC 1260

Qy       1553 CAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTT 1612
              |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       1261 CAATTGGCCCAACACTGTCTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTT 1320

Qy       1613 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1380

Qy       1673 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              |||||||||||||||||||||||||||||||||||
Db       1381 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1416
```

RESULT 6
AAS26566
ID AAS26566 standard; cDNA; 1129 BP.
XX
AC AAS26566;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 745.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

```
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   P-PSDB; AAU16579.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; SEQ ID NO 745; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence encodes a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

  Query Match             34.4%;  Score 588;  DB 4;  Length 1129;
  Best Local Similarity   99.2%;  Pred. No. 6.5e-285;
  Matches  988;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        103 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 162

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 222

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 282

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        283 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 342

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
              |||||||||||||||||||||||||||||||||||||| |||||| |||||||||| ||
Db        343 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGKTATCATYACTGGACCGCCARAG 402

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 462

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 522

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 582

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 642

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 702

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 762

Qy       1373 CCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1432
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 822

Qy       1433 AGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCC 1492
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        823 AGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCC 882

Qy       1493 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACC 1552
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        883 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACC 942

Qy       1553 CAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTT 1612
              |||||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||
Db        943 CAATTGGCCCAACACTGTCTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTT 1002

Qy       1613 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1003 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1062

Qy       1673 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||||||||||
Db        1063 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1098


RESULT 7
ABX73907
ID   ABX73907 standard; DNA; 1129 BP.
XX
AC   ABX73907;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polynucleotide #735.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
```

KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-00764864.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2003-147444/14.
DR P-PSDB; ABU55647.
XX
PT New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.
XX
PS Claim 1; SEQ ID NO 745; 402pp; English.
XX
CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC human novel polynucleotides of the invention
XX
SQ Sequence 1129 BP; 318 A; 296 C; 302 G; 210 T; 0 U; 3 Other;

Query Match 34.4%; Score 588; DB 7; Length 1129;
Best Local Similarity 99.2%; Pred. No. 6.5e-285;
Matches 988; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy       713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       103 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 162

Qy       773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       163 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 222

Qy       833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       223 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 282

Qy       893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       283 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 342

Qy       953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
            |||||||||||||||||||||||||||||||||||||| |||||| |||||||||||| ||
Db       343 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGKTATCATYACTGGACCGCCARAG 402

Qy      1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       403 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 462

Qy      1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       463 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 522

Qy      1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       523 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 582

Qy      1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       583 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 642

Qy      1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       643 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 702

Qy      1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1372
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       703 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 762

Qy      1373 CCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1432
            |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        763 CCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 822

Qy       1433 AGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCC 1492
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        823 AGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCC 882

Qy       1493 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACC 1552
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        883 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACC 942

Qy       1553 CAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTT 1612
              |||||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||
Db        943 CAATTGGCCCAACACTGTCTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTT 1002

Qy       1613 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1062

Qy       1673 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              |||||||||||||||||||||||||||||||||||
Db       1063 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1098
```

RESULT 8
AAK91969
ID AAK91969 standard; cDNA; 833 BP.
XX
AC AAK91969;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA 5'-end sequence, SEQ ID NO: 429.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
PS Claim 2; SEQ ID NO 429; 1380pp + Sequence Listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesising the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is the nucleotide sequence of the 5'-end of
CC a cDNA provided in the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in CD-
CC ROM format directly from EPO
XX
SQ Sequence 833 BP; 223 A; 225 C; 215 G; 166 T; 0 U; 4 Other;

```
  Query Match           24.2%;  Score 414;  DB 4;  Length 833;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-197;
  Matches  414;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        299 ACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAA 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAA 60

Qy        359 AAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTG 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTG 120

Qy        419 GAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATC 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATC 180

Qy        479 CTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAG 538
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAG 240

Qy        539 GTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTAC 598
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTAC 300

Qy        599 AACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAG 658
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAG 360

Qy        659 GAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 712
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 414
```

RESULT 9
AAK93655
ID AAK93655 standard; cDNA; 833 BP.
XX
AC AAK93655;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA clone representative sequence, SEQ ID NO: 2115.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
PS Example 11; SEQ ID NO 2115; 1380pp + Sequence Listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have

```
CC   been determined. Primers for synthesising the full length cDNA are useful
CC   for clarifying the function of the protein encoded by the cDNA. The full
CC   length clones were obtained by construction of full length enriched cDNA
CC   libraries that were synthesised by the oligo-capping method. The primers
CC   enable the production of the full length cDNA easily without any special
CC   methods. The present sequence was used as the representative sequence
CC   from a human clone which was used in homology searches to identify the
CC   clone. Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in CD-ROM format directly from
CC   EPO
XX
SQ   Sequence 833 BP; 223 A; 225 C; 215 G; 166 T; 0 U; 4 Other;

  Query Match             24.2%;  Score 414;  DB 4;  Length 833;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-197;
  Matches  414;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        299 ACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAA 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAA 60

Qy        359 AAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTG 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTG 120

Qy        419 GAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATC 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATC 180

Qy        479 CTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAG 538
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAG 240

Qy        539 GTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTAC 598
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTAC 300

Qy        599 AACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAG 658
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAG 360

Qy        659 GAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 712
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 414


RESULT 10
ABN42871
ID   ABN42871 standard; DNA; 60 BP.
XX
AC   ABN42871;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human spliced transcript detection oligonucleotide SEQ ID NO:15619.
XX
KW   Human; mouse; rat; splice transcript; detection; RNA transcript;
KW   splice variant; transcriptome; oligonucleotide library; ss.
XX
OS   Homo sapiens.
XX
PN   WO200210449-A2.
XX
PD   07-FEB-2002.
XX
PF   20-JUL-2001; 2001WO-IB001903.
XX
PR   28-JUL-2000; 2000US-0221607P.
PR   02-MAY-2001; 2001US-0287724P.
XX
PA   (COMP-) COMPUGEN INC.
XX
PI   Shoshan A,  Wasserman A,  Mintz E,  Mintz L,  Faigler S;
XX
DR   WPI; 2002-257383/30.
XX
PT   New oligonucleotide libraries comprising oligonucleotides which
PT   selectively hybridize to mRNAs transcribed from a transcription unit of a
PT   genome, useful for detecting tissue-, pathology-, and developmental-
PT   specific genes.
XX
PS   Example 1; SEQ ID NO 15619; 47pp; English.
XX
CC   The present invention describes oligonucleotide libraries for detecting
CC   messenger RNAs that populate a (sub-)transcriptome, where the (sub-
CC   )transcriptome comprises messenger RNAs transcribed from multiple
CC   transcription units that populate a genome. The library comprises several
CC   oligonucleotides, each capable of hybridising selectively to a set of
CC   messenger RNAs transcribed from a given transcription unit of the genome,
CC   which encodes one or more messenger RNA splice variants. The
CC   oligonucleotide libraries are useful for detecting mRNAs from a
CC   biological sample, in expression profiling studies, in qualitatively or
CC   quantitatively characterising the corresponding transcriptome, and in
CC   detecting RNA transcripts and splice variants of human or animal
CC   transcriptomes. The libraries may also be used as specialised mini
CC   libraries to detect transcripts of a sub-transcriptome under a particular
CC   biological or pathological state, and so allowing the detection of tissue
CC   - and pathology-specific genes such as those genes only expressed in
CC   specific tissue under a specific pathological condition; to detect
CC   developmental specific genes; and to detect RNA transcripts and splice
CC   variants of a transcriptome of a patient suffering from a particular
CC   disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from
CC   rats, humans and mice, which are used in the exemplification of the
CC   present invention. N.B. The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 60 BP; 17 A; 16 C; 14 G; 13 T; 0 U; 0 Other;

  Query Match             3.5%;  Score 60;  DB 6;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-19;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1243 AATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGC 1302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGC 60


RESULT 11
ABN59125
ID   ABN59125 standard; DNA; 60 BP.
XX
AC   ABN59125;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human spliced transcript detection oligonucleotide SEQ ID NO:31873.
XX
KW   Human; mouse; rat; splice transcript; detection; RNA transcript;
KW   splice variant; transcriptome; oligonucleotide library; ss.
XX
OS   Homo sapiens.
XX
PN   WO200210449-A2.
XX
PD   07-FEB-2002.
XX
PF   20-JUL-2001; 2001WO-IB001903.
XX
PR   28-JUL-2000; 2000US-0221607P.
PR   02-MAY-2001; 2001US-0287724P.
XX
PA   (COMP-) COMPUGEN INC.
XX
PI   Shoshan A,  Wasserman A,  Mintz E,  Mintz L,  Faigler S;
XX
```

DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of a
PT genome, useful for detecting tissue-, pathology-, and developmental-
PT specific genes.
XX
PS Example 1; SEQ ID NO 31873; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-
CC)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises several
CC oligonucleotides, each capable of hybridising selectively to a set of
CC messenger RNAs transcribed from a given transcription unit of the genome,
CC which encodes one or more messenger RNA splice variants. The
CC oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a particular
CC biological or pathological state, and so allowing the detection of tissue
CC - and pathology-specific genes such as those genes only expressed in
CC specific tissue under a specific pathological condition; to detect
CC developmental specific genes; and to detect RNA transcripts and splice
CC variants of a transcriptome of a patient suffering from a particular
CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from
CC rats, humans and mice, which are used in the exemplification of the
CC present invention. N.B. The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 60 BP; 15 A; 12 C; 18 G; 15 T; 0 U; 0 Other;

Query Match 3.5%; Score 60; DB 6; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.3e-19;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1470 CGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGA 1529
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGA 60

RESULT 12
ABN58997
ID ABN58997 standard; DNA; 60 BP.
XX
AC ABN58997;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:31745.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Homo sapiens.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-IB001903.
XX
PR 28-JUL-2000; 2000US-0221607P.
PR 02-MAY-2001; 2001US-0287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of a
PT genome, useful for detecting tissue-, pathology-, and developmental-
PT specific genes.
XX
PS Example 1; SEQ ID NO 31745; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-
CC)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises several
CC oligonucleotides, each capable of hybridising selectively to a set of
CC messenger RNAs transcribed from a given transcription unit of the genome,
CC which encodes one or more messenger RNA splice variants. The
CC oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a particular
CC biological or pathological state, and so allowing the detection of tissue
CC - and pathology-specific genes such as those genes only expressed in
CC specific tissue under a specific pathological condition; to detect
CC developmental specific genes; and to detect RNA transcripts and splice
CC variants of a transcriptome of a patient suffering from a particular
CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from
CC rats, humans and mice, which are used in the exemplification of the
CC present invention. N.B. The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 60 BP; 15 A; 12 C; 18 G; 15 T; 0 U; 0 Other;

Query Match 3.5%; Score 60; DB 6; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.3e-19;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1470 CGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGA 1529
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGA 60

RESULT 13
ABZ02948
ID ABZ02948 standard; DNA; 50 BP.
XX
AC ABZ02948;
XX
DT 09-JAN-2003 (first entry)
XX
DE Human leukocyte gene expression profiling probe SEQ ID NO 2939.
XX
KW T7; leukocyte; gene expression profiling; allograft rejection;
KW atherosclerosis; congestive heart failure; systemic lupus erythematosus;
KW rheumatoid arthritis; osteoarthritis; cytomegalovirus; infection; probe;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200257414-A2.
XX
PD 25-JUL-2002.
XX
PF 22-OCT-2001; 2001WO-US047856.
XX
PR 20-OCT-2000; 2000US-0241994P.
PR 08-JUN-2001; 2001US-0296764P.
XX
PA (BIOC-) BIOCARDIA INC.
XX
PI Wohlgemuth J, Fry K, Matcuk G, Altman P, Prentice J, Phillips J;
PI Ly N, Woodward R, Quertermous T, Johnson F;

XX
DR WPI; 2002-636525/68.
XX
PT New system for leukocyte expression profiling, diagnosing a disease, or
PT monitoring (the rate of) progression of a disease, e.g. atherosclerosis
PT or congestive heart failure, comprises diagnostic oligonucleotides.
XX
PS Claim 1; Page 421; 0pp; English.
XX
CC The invention relates to a system for detecting gene expression, which
CC comprises one or two isolated DNA molecules that detect expression of a
CC gene, where the gene corresponds to any of 8143 oligonucleotides
CC (ABZ00010-ABZ08152) each having 50 base pairs (bp). The system is useful
CC for leukocyte expression profiling. It is particularly useful for
CC diagnosing a disease, monitoring (rate of) progression of a disease,
CC predicting therapeutic outcome, determining prognosis for a patient,
CC predicting disease complications in an individual or monitoring response
CC to treatment in an individual. The diseases include cardiac allograft
CC rejection, kidney allograft rejection, liver allograft rejection,
CC atherosclerosis, congestive heart failure, systemic lupus erythematosus,
CC rheumatoid arthritis, osteoarthritis or cytomegalovirus infection
XX
SQ Sequence 50 BP; 18 A; 8 C; 13 G; 11 T; 0 U; 0 Other;

Query Match 2.9%; Score 50; DB 6; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1652 AGAGGGTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGT 1701
||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 AGAGGGTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGT 50

RESULT 14
AAZ10617
ID AAZ10617 standard; cDNA; 2224 BP.
XX
AC AAZ10617;
XX
DT 17-NOV-1999 (first entry)
XX
DE cDNA encoding a murine c-myc coding region determinant binding protein.
XX
KW c-myc coding region determinant binding protein; CRD-BP; tumor; c-myc;
KW endonucleolytic attack; half-life; breast cancer; colon cancer;
KW pancreatic cancer; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 131. .1864
FT /*tag= a
XX
PN WO9946594-A2.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004897.
XX
PR 09-MAR-1998; 98US-0077372P.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Ross J;
XX
DR WPI; 1999-551506/46.
DR P-PSDB; AAY30649.
XX
PT Diagnosing presence or absence of a tumor in a human by examining c-myc
PT coding region determinant-binding protein.
XX
PS Example; Fig 1A-D; 79pp; English.
XX
CC The present sequence encodes a murine c-myc coding region determinant
CC binding protein (CRD-BP). The presence or absence of a tumor can be
CC determined by determining the levels of CRD-BP present in the suspect
CC tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack
CC and so prolongs its half-life. The methods are used for diagnosing
CC presence or absence of a tumor in a human, especially breast, colon and
CC pancreatic cancer. They are also used to inhibit cancer cell growth
XX
SQ Sequence 2224 BP; 580 A; 609 C; 618 G; 417 T; 0 U; 0 Other;

Query Match 1.8%; Score 31; DB 2; Length 2224;
Best Local Similarity 100.0%; Pred. No. 9.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 TGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
|||||||||||||||||||||||||||||||
Db 561 TGGAGAACCATGCCCTGAAGGTCTCCTACAT 591

RESULT 15
AAZ10625/c
ID AAZ10625 standard; DNA; 24 BP.
XX
AC AAZ10625;
XX
DT 17-NOV-1999 (first entry)
XX
DE PCR primer used to amplify murine CRD-BP cDNA.
XX
KW c-myc coding region determinant binding protein; CRD-BP; tumor; c-myc;
KW endonucleolytic attack; half-life; breast cancer; colon cancer;
KW pancreatic cancer; PCR primer; ss.
XX
OS Synthetic.
OS Mus musculus.
XX
PN WO9946594-A2.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004897.
XX
PR 09-MAR-1998; 98US-0077372P.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Ross J;
XX
DR WPI; 1999-551506/46.
XX
PT Diagnosing presence or absence of a tumor in a human by examining c-myc
PT coding region determinant-binding protein.
XX
PS Example; Page 27; 79pp; English.
XX
CC The present PCR primer was used to amplify cDNA encoding murine c-myc
CC coding region determinant binding protein (CRD-BP). The presence or
CC absence of a tumor can be determined by determining the levels of CRD-BP
CC present in the suspect tissue, where the CRD-BP shields c-myc RNA from
CC endonucleolytic attack and so prolongs its half-life. The methods are
CC used for diagnosing presence or absence of a tumor in a human, especially
CC breast, colon and pancreatic cancer. They are also used to inhibit cancer
CC cell growth
XX
SQ Sequence 24 BP; 3 A; 8 C; 5 G; 8 T; 0 U; 0 Other;

Query Match 1.4%; Score 24; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 508 CATTGGCAAGGAAGGACGGAACCT 531
||||||||||||||||||||||||
Db 24 CATTGGCAAGGAAGGACGGAACCT 1

Search completed: July 13, 2004, 23:52:48
Job time : 704 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 23:20:54 ; Search time 4590 Seconds
(without alignments)
11112.114 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 1708
Sequence: 1 agggacgctgccgcaccgcc.....atttccttcaggttttaaaa 1708

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	676	39.6	1042	12	BM928196	BM928196 AGENCOURT
	2	657	38.5	1085	13	BU190334	BU190334 AGENCOURT
	3	606	35.5	955	13	BQ651665	BQ651665 AGENCOURT
	4	584	34.2	950	13	BQ648383	BQ648383 AGENCOURT
	5	545	31.9	983	13	BQ643920	BQ643920 AGENCOURT
	6	526	30.8	831	12	BG115319	BG115319 602316274
c	7	515	30.2	919	13	BX327672	BX327672 BX327672
	8	499	29.2	896	10	BF984962	BF984962 602308484
	9	474	27.8	1042	12	BM561057	BM561057 AGENCOURT
	10	426	24.9	1061	13	BQ647561	BQ647561 AGENCOURT
	11	419	24.5	712	13	BQ225582	BQ225582 AGENCOURT
	12	417	24.4	953	13	BQ647360	BQ647360 AGENCOURT
	13	382	22.4	884	13	BU598549	BU598549 AGENCOURT
c	14	295	17.3	761	9	AW003366	AW003366 wq65h07.x
	15	291	17.0	568	14	CB146278	CB146278 K-EST0201
	16	290	17.0	545	13	BX089485	BX089485 BX089485
	17	288	16.9	1173	12	BM467135	BM467135 AGENCOURT
	18	272	15.9	982	13	BQ652180	BQ652180 AGENCOURT
	19	268	15.7	978	12	BG115593	BG115593 602317291
	20	246	14.4	764	12	BM006191	BM006191 603614013
	21	232	13.6	929	13	BQ652586	BQ652586 AGENCOURT
	22	231	13.5	486	9	AL704124	AL704124 DKFZp686B
c	23	231	13.5	891	13	BX372273	BX372273 BX372273
c	24	219	12.8	484	9	AI337147	AI337147 qx83c08.x
c	25	213	12.5	550	9	AA978341	AA978341 oq40a04.s
	26	194	11.4	896	13	BQ647425	BQ647425 AGENCOURT
c	27	172	10.1	517	10	BE550506	BE550506 7a27f01.x
c	28	167	9.8	1201	13	BX401619	BX401619 BX401619
c	29	158	9.3	656	10	AW303505	AW303505 xv19h09.x
c	30	151	8.8	478	9	AA196977	AA196977 zq60h02.s
c	31	138	8.1	257	10	BE466868	BE466868 hz28a05.x
	32	135	7.9	249	9	AA196774	AA196774 zq60f02.r
c	33	131	7.7	1201	13	BX333010	BX333010 BX333010
c	34	121	7.1	393	28	AQ096481	AQ096481 HS_3029_B
	35	121	7.1	628	29	AG064887	AG064887 Pan trogl
c	36	113	6.6	265	9	AI962664	AI962664 wq42f07.x
c	37	113	6.6	317	10	AW206272	AW206272 UI-H-BI1-
c	38	113	6.6	461	9	AA196759	AA196759 zq60f02.s
c	39	110	6.4	258	9	AI989990	AI989990 ws28f09.x
c	40	109	6.4	257	9	AI335528	AI335528 qt24b06.x
c	41	109	6.4	257	9	AI671622	AI671622 wb41h04.x
c	42	109	6.4	257	10	BE466712	BE466712 hz24e03.x
c	43	109	6.4	258	10	AW196265	AW196265 xm31e06.x
c	44	109	6.4	258	10	BE550313	BE550313 7a21h08.x
c	45	109	6.4	369	10	BF115470	BF115470 7m89a03.x

ALIGNMENTS

RESULT 1
BM928196
LOCUS BM928196 1042 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6715416 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5797961 5', mRNA sequence.
ACCESSION BM928196
VERSION BM928196.1 GI:19378575
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1042)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2025 row: c column: 18
High quality sequence stop: 662.

```
FEATURES             Location/Qualifiers
     source          1. .1042
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5797961"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match             39.6%;  Score 676;  DB 12;  Length 1042;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  676;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 71

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 131

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 191

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 251

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 311

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 371

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 431

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 491

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 551

Qy        577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 611

Qy        637 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 671

Qy        697 TGTGGCTGCCATGAGC 712
              ||||||||||||||||
Db        672 TGTGGCTGCCATGAGC 687
```

```
RESULT 2
BU190334
LOCUS       BU190334                1085 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7186901 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6001834
            5', mRNA sequence.
ACCESSION   BU190334
VERSION     BU190334.1  GI:22704318
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1085)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2269  row: b  column: 11
            High quality sequence stop: 671.
FEATURES             Location/Qualifiers
     source          1. .1085
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6001834"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match             38.5%;  Score 657;  DB 13;  Length 1085;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  657;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         56 GGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCA 115
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         32 GGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCA 91

Qy        116 CAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGC 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         92 CAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGC 151

Qy        176 TCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTC 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        152 TCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTC 211

Qy        236 CTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAAC 295
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        212 CTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAAC 271

Qy        296 ATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCT 355
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        272 ATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCT 331

Qy        356 GAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATC 415
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        332 GAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATC 391

Qy        416 TTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAG 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        392 TTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAG 451

Qy        476 ATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAG 535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        452 ATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAG 511

Qy        536 AAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTT 595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        512 AAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTT 571

Qy        596 TACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAG 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 TACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAG 631

Qy        656 CAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        632 CAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 688
```

RESULT 3
BQ651665
LOCUS BQ651665 955 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8297761 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269842 5', mRNA sequence.
ACCESSION BQ651665
VERSION BQ651665.1 GI:21775837
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 955)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2444 row: a column: 11
High quality sequence stop: 588.
FEATURES Location/Qualifiers
source 1. .955
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6269842"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN

Query Match 35.5%; Score 606; DB 13; Length 955;
Best Local Similarity 100.0%; Pred. No. 2.7e-301;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 71

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 131

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 191

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 251

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 311

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 371

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 431

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 491

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 551

Qy        577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 611

Qy        637 GAATTG 642
              ||||||
Db        612 GAATTG 617
```

RESULT 4
BQ648383
LOCUS BQ648383 950 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8342265 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6268202 5', mRNA sequence.
ACCESSION BQ648383
VERSION BQ648383.1 GI:21772555
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 950)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2439 row: m column: 03
High quality sequence stop: 599.

```
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6268202"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match           34.2%;  Score 584;  DB 13;  Length 950;
  Best Local Similarity  100.0%;  Pred. No. 6.3e-290;
  Matches  584;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 71

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 131

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 191

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 251

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 311

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 371

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 431

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 491

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 551

Qy        577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTG 620
              |||||||||||||||||||||||||||||||||||||||||||
Db        552 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTG 595


RESULT 5
BQ643920
LOCUS       BQ643920                983 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8303966 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269413
            5', mRNA sequence.
ACCESSION   BQ643920
VERSION     BQ643920.1  GI:21768092
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 983)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2442  row: o  column: 14
            High quality sequence stop: 491.
FEATURES             Location/Qualifiers
     source          1. .983
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6269413"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match           31.9%;  Score 545;  DB 13;  Length 983;
  Best Local Similarity  99.8%;  Pred. No. 9.3e-270;
  Matches  595;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 71

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 131

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 191

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 251

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 311

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 371

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 431

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        432 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 491

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        492 GGAAGGACGGAACCTGAAGAAAGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 551

Qy        577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCA 632
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCA 607


RESULT 6
BG115319
LOCUS       BG115319                831 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  602316274F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4416354 5',
            mRNA sequence.
ACCESSION   BG115319
VERSION     BG115319.1  GI:12608825
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 831)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10147  row: d  column: 19
            High quality sequence stop: 731.
FEATURES             Location/Qualifiers
     source          1. .831
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4416354"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_88"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             30.8%;  Score 526;  DB 12;  Length 831;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-260;
  Matches  526;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        820 CTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGT 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGT 60

Qy        880 GGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGC 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGC 120

Qy        940 CTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCAC 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCAC 180

Qy       1000 TGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGA 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGA 240

Qy       1060 GAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATC 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATC 300

Qy       1120 AGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGAC 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGAC 360

Qy       1180 GGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGT 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGT 420

Qy       1240 GAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCT 480

Qy       1300 GGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCC 1345
              |||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCC 526


RESULT 7
BX327672/c
LOCUS       BX327672                919 bp    mRNA    linear   EST 02-MAY-2003
DEFINITION  BX327672 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC018YI18 5-PRIME, mRNA sequence.
ACCESSION   BX327672
VERSION     BX327672.1  GI:30342696
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 919)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 8684.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0BAG007ZH09_CS00648_1&cluster=8684.r.
            Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0BAG007ZH09_CS00648_1.
FEATURES             Location/Qualifiers
     source          1. .919
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC018YI18"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             30.2%;  Score 515;  DB 13;  Length 919;
  Best Local Similarity   99.8%;  Pred. No. 2.9e-254;
  Matches  565;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 556

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        555 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 496

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        495 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 436

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 376

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 316

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
                | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 GAAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 256

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 196

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 136

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 76

Qy        577 GTTGCAAGACCTTACCCTTTACAACC 602
                ||||||||||||||||||||||||||
Db         75 GTTGCAAGACCTTACCCTTTACAACC 50
```

RESULT 8
BF984962
LOCUS BF984962 896 bp mRNA linear EST 23-JAN-2001
DEFINITION 602308484F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4399556 5', mRNA sequence.
ACCESSION BF984962
VERSION BF984962.1 GI:12387774
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 896)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10103 row: h column: 21
High quality sequence stop: 652.
FEATURES Location/Qualifiers
source 1. .896
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4399556"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_88"
/note="Organ: small intestine; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
ORIGIN

```
  Query Match             29.2%;  Score 499;  DB 10;  Length 896;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-246;
  Matches  499;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 196

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 256

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 316

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 376

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 436

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 496

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        497 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 556

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        557 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 616

Qy       1193 GTAGTACCAAGAGACCAGA 1211
                ||||||||||||||||||
Db        617 GTAGTACCAAGAGACCAGA 635
```

RESULT 9
BM561057
LOCUS BM561057 1042 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6566228 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5550887 5', mRNA sequence.
ACCESSION BM561057
VERSION BM561057.1 GI:18805998
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1042)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12263 row: l column: 24

```
          High quality sequence stop: 682.
FEATURES             Location/Qualifiers
     source          1. .1042
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5550887"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.   Library constructed by Life
                     Technologies."
ORIGIN

  Query Match             27.8%;  Score 474;  DB 12;  Length 1042;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-233;
  Matches  474;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        283 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 342

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 402

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 462

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 522

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 582

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 642

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 702

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCT 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCT 756


RESULT 10
BQ647561
LOCUS       BQ647561               1061 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8349614 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6284767
            5', mRNA sequence.
ACCESSION   BQ647561
VERSION     BQ647561.1  GI:21771733
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1061)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2482  row: o  column: 08
            High quality sequence stop: 443.
FEATURES             Location/Qualifiers
     source          1. .1061
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6284767"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match             24.9%;  Score 426;  DB 13;  Length 1061;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-208;
  Matches  426;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 71

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 131

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 191

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 251

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 311

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 371

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 431

Qy        457 TGACGA 462
              ||||||
Db        432 TGACGA 437


RESULT 11
BQ225582
LOCUS       BQ225582               712 bp    mRNA    linear   EST 02-MAY-2002
DEFINITION  AGENCOURT_7593363 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020837
            5', mRNA sequence.
ACCESSION   BQ225582
VERSION     BQ225582.1  GI:20406982
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13225  row: j  column: 06
            High quality sequence stop: 625.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6020837"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_70"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies."
ORIGIN

  Query Match            24.5%;  Score 419;  DB 13;  Length 712;
  Best Local Similarity  99.3%;  Pred. No. 1.2e-204;
  Matches  669;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1031 GGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAG 1090
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db             1 GGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAG 60

Qy       1091 CTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGA 1150
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            61 CTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGA 120

Qy       1151 AAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAG 1210
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           121 AAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAG 180

Qy       1211 ACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAG 1270
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           181 ACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAG 240

Qy       1271 ATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGA 1330
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           241 ATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGA 300

Qy       1331 CAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCC 1390
                ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db           301 CAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCCCTTCGAGTCC 360

Qy       1391 AGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGG 1450
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           361 AGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGG 420

Qy       1451 GAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATG 1510
                |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db           421 GAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATG 480

Qy       1511 TTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGCCCAACACTGT 1570
                |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db           481 TTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACCCAATTGGCCCAACACTGT 540

Qy       1571 NTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGGATTGTTTAAA 1630
                |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db           541 CTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTTAAACGTGGATTGTTTAAA 600

Qy       1631 GAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGAAAAATAAAAT 1690
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           601 GAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGAAAAATAAAAT 660

Qy       1691 TTCCTTCAGGTTTT 1704
                ||||||||||||||
Db           661 TTCCTTCAGGTTTT 674


RESULT 12
BQ647360
LOCUS       BQ647360                953 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8443628 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6284231
            5', mRNA sequence.
ACCESSION   BQ647360
VERSION     BQ647360.1  GI:21771532
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 953)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2481  row: h  column: 24
            High quality sequence stop: 651.
FEATURES             Location/Qualifiers
     source          1. .953
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6284231"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match            24.4%;  Score 417;  DB 13;  Length 953;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-203;
  Matches  417;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 364

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 424

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 484

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 544

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 604

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 664

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAAC 453
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        665 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAAC 721

RESULT 13
BU598549
LOCUS       BU598549                 884 bp    mRNA    linear   EST 20-SEP-2002
DEFINITION  AGENCOURT_8822656 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6457432
            5', mRNA sequence.
ACCESSION   BU598549
VERSION     BU598549.1  GI:23250308
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: NCI
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2639  row: e  column: 17
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1..884
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6457432"
                     /tissue_type="mixed (pool of 40 RNAs)"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_142"
                     /note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc);
                     Site_2: SfiI (ggccgcctcggcc); Double-stranded cDNA was
                     prepared from a pool of 40 cell line polyA+ RNAs (bladder
                     - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
                     4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
                     kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
                     ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
                     gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
                     used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the >0.5 kb
                     size fraction (other fractions present in NIH_MGC_141).
                     Library created in the laboratory of M. Brownstein (NIMH,
                     NIH). Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             22.4%;  Score 382;  DB 13;  Length 884;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-185;
  Matches  382;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        172 GGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCG 231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCG 62

Qy        232 GCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 GCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCG 122

Qy        292 CAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGC 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGC 182

Qy        352 AGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGAT 411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 AGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGAT 242

Qy        412 GATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCT 471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 GATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCT 302

Qy        472 GAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCT 531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCT 362

Qy        532 GAAGAAGGTAGAGCAAGATACC 553
              ||||||||||||||||||||||
Db        363 GAAGAAGGTAGAGCAAGATACC 384

RESULT 14
AW003366/c
LOCUS       AW003366                 761 bp    mRNA    linear   EST 08-MAR-2000
DEFINITION  wq65h07.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2476189 3'
            similar to TR:O88477 O88477 CODING REGION DETERMINANT BINDING
            PROTEIN. ;, mRNA sequence.
ACCESSION   AW003366
VERSION     AW003366.1  GI:5850282
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 761)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1368   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 456.
FEATURES             Location/Qualifiers
     source          1..761
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2476189"
                     /tissue_type="pooled germ cell tumors"
```

/lab_host="DH10B"
/clone_lib="NCI_CGAP_GC6"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 17.3%; Score 295; DB 9; Length 761;
Best Local Similarity 100.0%; Pred. No. 1.5e-140;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1024 GGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGA 1083
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        611 GGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGA 552

Qy       1084 AGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAA 1143
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         551 AGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGGCAA 492

Qy       1144 AGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAG 1203
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         491 AGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACCAAG 432

Qy       1204 AGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGC 1263
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         431 AGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTATGC 372

Qy       1264 CAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAG 1318
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         371 CAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAG 317
```

RESULT 15
CB146278
LOCUS CB146278 568 bp mRNA linear EST 29-JAN-2003
DEFINITION K-EST0201433 L16HLK3 Homo sapiens cDNA clone L16HLK3-20-H04 5', mRNA sequence.
ACCESSION CB146278
VERSION CB146278.1 GI:28126979
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 568)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 20 row: H column: 04
High quality sequence stop: 568.
FEATURES Location/Qualifiers
source 1. .568
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L16HLK3-20-H04"
/cell_line="HLK-3"
/lab_host="Top10F'"
/clone_lib="L16HLK3"
/note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI; The library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."

ORIGIN

Query Match 17.0%; Score 291; DB 14; Length 568;
Best Local Similarity 100.0%; Pred. No. 1.6e-138;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        278 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 337

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        338 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 397

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 457

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        458 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 517

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGA 1003
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGA 568
```

Search completed: July 14, 2004, 03:02:41
Job time : 4594 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 13, 2004, 12:24:58 ; Search time 30 Seconds
(without alignments)
10953.008 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 560
Sequence: 1 agggacgctgccgcaccgcc.....atttccttcaggttttaaaa 1708

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565818

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09270437/runat_13072004_121956_9997/app_query.fasta_1.1863
-DB=PIR_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437_@CGN_1_1_44_@runat_13072004_121956_9997 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	8	1.4	95	2	H95092	hypothetical prote
c	2	8	1.4	95	2	E97960	hypothetical prote
c	3	8	1.4	185	2	I39736	hypothetical prote
c	4	8	1.4	217	2	AD2683	Conserved hypothet
c	5	8	1.4	221	2	AF3195	transcription regu
	6	8	1.4	223	2	A83859	menaquinol-cytochr
	7	8	1.4	224	2	I39943	menaquinol-cytochr
c	8	8	1.4	227	2	A70036	capsular polysacch
	9	8	1.4	247	2	B90733	probable major tai
	10	8	1.4	249	2	D85583	probable tail comp
c	11	8	1.4	253	2	B97465	hypothetical prote
c	12	8	1.4	259	2	G64831	probable membrane
c	13	8	1.4	259	2	C90754	hypothetical prote
c	14	8	1.4	259	2	A85618	hypothetical prote
	15	8	1.4	301	2	AF2223	heterodisulfide re
	16	8	1.4	306	2	AF0876	agmatine ureohydro
	17	8	1.4	340	2	AB2584	hypothetical prote
	18	8	1.4	340	2	B97366	frcB protein (AF19
	19	8	1.4	351	2	AC3090	conserved hypothet
	20	8	1.4	351	2	G98196	hypothetical prote
	21	8	1.4	357	2	B81396	probable aminotran
c	22	8	1.4	367	2	T36278	hypothetical prote
	23	8	1.4	374	2	T07887	G box-binding prot
	24	8	1.4	376	2	T07882	G box-binding prot
	25	8	1.4	457	2	B87269	hypothetical prote
	26	8	1.4	475	2	G70861	hypothetical prote
	27	8	1.4	532	2	E95949	probable oligopept
c	28	8	1.4	546	2	AI0278	probable AMP nucle
	29	8	1.4	555	2	T30349	structural protein
	30	8	1.4	573	2	C85433	splicing factor-li
	31	8	1.4	573	2	C86266	F3F19.21 protein -
	32	8	1.4	600	2	T09676	probable ATP-depen
	33	8	1.4	638	2	F75547	anthranilate synth
c	34	8	1.4	645	2	H96011	asparagine synthas
	35	8	1.4	646	1	S15901	chromogranin B pre
	36	8	1.4	680	2	T25832	hypothetical prote
c	37	8	1.4	735	2	A84829	hypothetical prote
c	38	8	1.4	755	2	T47731	hypothetical prote
c	39	8	1.4	759	2	T39090	probable integral
	40	8	1.4	812	2	T34180	hypothetical prote
c	41	8	1.4	963	2	T26022	hypothetical prote
	42	8	1.4	1030	2	S73460	probable helicase
	43	8	1.4	1198	2	B88279	protein C08H9.2 [i
	44	8	1.4	1220	2	T19117	hypothetical prote
	45	8	1.4	1269	2	F88101	protein W09G10.4 [

ALIGNMENTS

RESULT 1
H95092
hypothetical protein SP0800 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: H95092
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: H95092
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-95 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK74937.1; PID:g14972276; GSPDB:GN00164; TIGR:SP
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0800

Alignment Scores:

Pred. No.:	63.8	Length:	95
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.44%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x H95092 (1-95)

```
Qy       1702 AACCTGAAGGAAATTTTATTTTTC 1679
              |||||||||||||||||||||||
Db         36 AsnLeuLysGluIleLeuPhePhe 43
```

RESULT 2
E97960

hypothetical protein spr0709 [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: E97960
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: E97960
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-95 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK99513.1; PID:g15458299; GSPDB:GN00174
C;Genetics:
A;Gene: spr0709

Alignment Scores:

Pred. No.:	63.8	Length:	95
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.44%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x E97960 (1-95)

```
Qy       1702 AACCTGAAGGAAATTTTATTTTTC 1679
                ||||||||||||||||||||||||
Db         36 AsnLeuLysGluIleLeuPhePhe 43
```

RESULT 3
I39736
hypothetical protein 6 - Anabaena variabilis
C;Species: Anabaena variabilis
C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 15-Oct-1999
C;Accession: S68185; I39736
R;Schmitz, O.; Boison, G.; Hilscher, R.; Hundeshagen, B.; Zimmer, W.; Lottspeich, F.; Bo
Eur. J. Biochem. 233, 266-276, 1995
A;Title: Molecular biological analysis of a bidirectional hydrogenase from cyanobacteria
A;Reference number: I39730; MUID:96061958; PMID:7588754
A;Accession: S68185
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-185 <SCH>
A;Cross-references: EMBL:X79285; NID:g1032475; PIDN:CAA55879.1; PID:g1032482
A;Experimental source: ATCC 29413
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994

Alignment Scores:

Pred. No.:	58	Length:	185
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.44%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x I39736 (1-185)

```
Qy        478 GATCTTCAGGGGAACCTCGTCAGC 455
                ||||||||||||||||||||||||
Db         90 AspLeuGlnGlyAsnLeuValSer 97
```

RESULT 4
AD2683
Conserved hypothetical protein Atu0868 [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AD2683
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AD2683
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-217 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL41882.1; PID:g17739245; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu0868
A;Map position: circular chromosome

Alignment Scores:

Pred. No.:	56.7	Length:	217
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.44%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x AD2683 (1-217)

```
Qy       1661 TCCACCCTCTTGGTGGGGCCTGGA 1638
                ||||||||||||||||||||||||
Db        103 SerThrLeuLeuValGlyProGly 110
```

RESULT 5
AF3195
transcription regulator, TetR family Atu5291 [imported] - Agrobacterium tumefaciens (str
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AF3195
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF3195
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-221 <KUR>
A;Cross-references: GB:AE008687; PIDN:AAL45980.1; PID:g17743733; GSPDB:GN00188
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu5291
A;Genome: plasmid

Alignment Scores:

Pred. No.:	56.6	Length:	221
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.44%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x AF3195 (1-221)

```
Qy        227 GGGGGATGTCCACTTGCTGCTGCT 204
                ||||||||||||||||||||||||
Db        111 GlyGlyCysProLeuAlaAlaAla 118
```

RESULT 6
AB3859
menaquinol-cytochrome c reductase (cytochrome b subunit) qcrB [imported] - Bacillus halo
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 03-Jun-2002

C;Accession: A83859
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: A83859
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-223 <STO>
A;Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BAB05392.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: qcrB

Alignment Scores:

Pred. No.:	56.5	Length:	223
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.43%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x A83859 (1-223)

```
Qy        786 CTCCCAGCAGCGTTACTGGGGCTG 809
              ||||||||||||||||||||||||
Db        199 LeuProAlaAlaLeuLeuGlyLeu 206
```

RESULT 7
I39943
menaquinol-cytochrome-c reductase (EC 1.10.2.-) cytochrome b6 qcrB - Bacillus stearother
C;Species: Bacillus stearothermophilus
C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 04-Mar-2000
C;Accession: I39943
R;Sone, N.; Sawa, G.; Sone, T.; Noguchi, S.
J. Biol. Chem. 2, 10612-10617, 1995
A;Title: Thermophilic bacilli have split cytochrome b genes for cytochrome b6 and subuni
A;Reference number: I39943
A;Accession: I39943
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-224 <RES>
A;Cross-references: GB:D45410; NID:g902924; PID:g643677
C;Genetics:
A;Gene: petB
A;Start codon: GTG
C;Superfamily: cytochrome b6; cytochrome b6 homology
C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; oxidoreductase
F;16-224/Domain: cytochrome b6 homology <CB6>
F;43/Binding site: heme (Cys) (covalent) (probably high potential) #status predicted
F;94,196/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted
F;108,211/Binding site: heme iron (His) (axial ligands) (high potential) #status predict

Alignment Scores:

Pred. No.:	56.5	Length:	224
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.43%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x I39943 (1-224)

```
Qy        786 CTCCCAGCAGCGTTACTGGGGCTG 809
              ||||||||||||||||||||||||
Db        200 LeuProAlaAlaLeuLeuGlyLeu 207
```

RESULT 8
A70036
capsular polysaccharide biosynthesis homolog yveL - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: A70036
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, I
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: A70036
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-227 <KUN>
A;Cross-references: GB:Z99121; GB:AL009126; NID:g2635827; PIDN:CAB15441.1; PID:g2635949
A;Experimental source: strain 168
C;Genetics:
A;Gene: yveL
C;Superfamily: capsular polysaccharide biosynthesis protein cpsC

Alignment Scores:

Pred. No.:	56.4	Length:	227
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.44%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x A70036 (1-227)

```
Qy       1367 CTGGTCACTTCCTCCGTGCCTGGG 1344
              ||||||||||||||||||||||||
Db         48 LeuValThrSerSerValProGly 55
```

RESULT 9
B90733
probable major tail protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: B90733
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B90733
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-247 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34257.1; PID:g13360293; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0834
C;Superfamily: phage lambda major tail protein V

Alignment Scores:

Pred. No.:	55.7	Length:	247
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.43%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x B90733 (1-247)

```
Qy        359 AAAGCCATCAGTGTGCACTCCACC 382
              ||||||||||||||||||||||||
```

```
Db        193 LysAlaIleSerValHisSerThr 200

RESULT 10
D85583
probable tail component of prophage CP-933K Z0972 [imported] - Escherichia coli (strain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85583
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85583
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-249 <STO>
A;Cross-references: GB:AE005174; NID:g12513740; PIDN:AAG55128.1; GSPDB:GN00145; UWGP:Z09
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0972
C;Superfamily: phage lambda major tail protein V

Alignment Scores:
Pred. No.:                55.6          Length:          249
Score:                    8.00          Matches:         8
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              1.43%         Indels:          0
DB:                       2             Gaps:            0

US-09-270-437D-5 (1-1708) x D85583 (1-249)

Qy        359 AAAGCCATCAGTGTGCACTCCACC 382
              ||||||||||||||||||||||||
Db        195 LysAlaIleSerValHisSerThr 202

RESULT 11
B97465
hypothetical protein AGR_C_1585 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: B97465
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: B97465
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK86675.1; PID:g15155859; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_1585
A;Map position: circular chromosome

Alignment Scores:
Pred. No.:                55.5          Length:          253
Score:                    8.00          Matches:         8
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              1.44%         Indels:          0
DB:                       2             Gaps:            0

US-09-270-437D-5 (1-1708) x B97465 (1-253)

Qy       1661 TCCACCCTCTTGGTGGGGCCTGGA 1638
              ||||||||||||||||||||||||
Db        139 SerThrLeuLeuValGlyProGly 146

RESULT 12
G64831
probable membrane protein ycbC - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: G64831
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G64831
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-259 <BLAT>
A;Cross-references: GB:AE000194; GB:U00096; NID:g1787148; PIDN:AAC74006.1; PID:g1787150;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: ycbC
C;Superfamily: conserved hypothetical protein aq_1986
C;Keywords: transmembrane protein
F;13-29/Domain: transmembrane #status predicted <TM1>
F;39-55/Domain: transmembrane #status predicted <TM2>

Alignment Scores:
Pred. No.:                55.3          Length:          259
Score:                    8.00          Matches:         8
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              1.44%         Indels:          0
DB:                       2             Gaps:            0

US-09-270-437D-5 (1-1708) x G64831 (1-259)

Qy        479 GGATCTTCAGGGGAACCTCGTCAG 456
              ||||||||||||||||||||||||
Db        251 GlySerSerGlyGluProArgGln 258

RESULT 13
C90754
hypothetical protein ECs1003 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: C90754
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90754
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-259 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34426.1; PID:g13360462; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1003
C;Superfamily: conserved hypothetical protein aq_1986

Alignment Scores:
Pred. No.:                55.3          Length:          259
Score:                    8.00          Matches:         8
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              1.44%         Indels:          0
DB:                       2             Gaps:            0

US-09-270-437D-5 (1-1708) x C90754 (1-259)

Qy        479 GGATCTTCAGGGGAACCTCGTCAG 456
              ||||||||||||||||||||||||
Db        251 GlySerSerGlyGluProArgGln 258

RESULT 14
A85618
```

hypothetical protein ycbC [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: A85618
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A85618
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-259 <STO>
A;Cross-references: GB:AE005174; NID:g12514089; PIDN:AAG55405.1; GSPDB:GN00145; UWGP:Z12
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: ycbC
C;Superfamily: conserved hypothetical protein aq_1986

Alignment Scores:

Pred. No.:	55.3	Length:	259
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.44%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x A85618 (1-259)

```
Qy        479 GGATCTTCAGGGGAACCTCGTCAG 456
              ||||||||||||||||||||||||
Db        251 GlySerSerGlyGluProArgGln 258
```

RESULT 15
AF2223
heterodisulfide reductase, chain B [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AF2223
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AF2223
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-301 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75040.1; PID:g17132436; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all3341
C;Superfamily: Acidianus ambivalens succinate dehydrogenase chain C

Alignment Scores:

Pred. No.:	54.2	Length:	301
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.43%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x AF2223 (1-301)

```
Qy        144 GCTTTGGCTCTCGGGGTCAGCCCC 167
              ||||||||||||||||||||||||
Db        273 AlaLeuAlaLeuGlyValSerPro 280
```

Search completed: July 13, 2004, 12:35:45
Job time : 37 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 21:42:59 ; Search time 6722 Seconds
(without alignments)
11013.077 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 1708
Sequence: 1 agggacgctgccgcaccgcc.....atttccttcaggttttaaaa 1708

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1708	100.0	1708	6	AR171864	AR171864 Sequence
2		1708	100.0	1708	6	AR343073	AR343073 Sequence
3		1708	100.0	1708	6	BD209924	BD209924 Isolated
4		1672	97.9	1946	6	AR171866	AR171866 Sequence
5		1672	97.9	1946	6	AR343075	AR343075 Sequence
6		1672	97.9	1946	6	BD209926	BD209926 Isolated
7		997	58.4	2130	9	AF117106	AF117106 Homo sapi
8		821	48.1	2381	9	AF198254	AF198254 Homo sapi
9		741	43.4	2780	6	BD127811	BD127811 Primer fo
10		741	43.4	2780	9	AK074915	AK074915 Homo sapi
11		414	24.2	833	6	BD124998	BD124998 Primer fo
12		414	24.2	833	6	BD126684	BD126684 Primer fo
13	c	284	16.6	107848	9	AC105030	AC105030 Homo sapi
14		284	16.6	154604	2	AC104974	AC104974 Homo sapi
15	c	284	16.6	159122	2	AC025556	AC025556 Homo sapi
16	c	284	16.6	168613	9	AC091133	AC091133 Homo sapi
17	c	208	12.2	65435	2	AC104587	AC104587 Homo sapi
18		186	10.9	65435	2	AC104587	AC104587 Homo sapi
19		136	8.0	164950	2	AC105191	AC105191 Homo sapi
20	c	121	7.1	67462	2	AC091595	AC091595 Homo sapi
21		62	3.6	67462	2	AC091595	AC091595 Homo sapi
22		38	2.2	216616	10	AC094527	AC094527 Rattus no
23	c	38	2.2	220434	2	AC106496	AC106496 Rattus no
24		38	2.2	232468	2	AC094487	AC094487 Rattus no
25	c	36	2.1	662	9	HSA331847	AJ331847 Homo sapi
26	c	36	2.1	845	9	HSA334465	AJ334465 Homo sapi
27		32	1.9	2105	10	AF541940	AF541940 Rattus no
28		32	1.9	196392	2	AC120322	AC120322 Rattus no
29		31	1.8	2223	10	AF061569	AF061569 Mus muscu
30		31	1.8	2224	6	AR160244	AR160244 Sequence
31		31	1.8	2444	10	BC051679	BC051679 Mus muscu
32	c	31	1.8	71109	10	AL606704	AL606704 Mouse DNA
33		31	1.8	188834	10	AC098642	AC098642 Genomic s
34	c	31	1.8	239837	10	AC084407	AC084407 Mus Muscu
35	c	26	1.5	187119	10	AC123533	AC123533 Mus muscu
36		26	1.5	219696	2	AC099577	AC099577 Mus muscu
37	c	25	1.5	177384	2	AC091786	AC091786 Mus muscu
38	c	24	1.4	24	6	AR160252	AR160252 Sequence
39		23	1.3	23	6	AR160250	AR160250 Sequence
40	c	23	1.3	1493	10	AF076930	AF076930 Mus muscu
41	c	23	1.3	178165	2	AC109259	AC109259 Mus muscu
42		23	1.3	180104	9	DJ526N18	AF123462 Homo sapi
43		22	1.3	1536	6	E00532	E00532 DNA encodin
44		22	1.3	1787	8	GMHSP2	X01104 Soybean hea
45	c	22	1.3	170348	2	AC105168	AC105168 Mus muscu

ALIGNMENTS

RESULT 1
AR171864
LOCUS AR171864 1708 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 5 from patent US 6297364.
ACCESSION AR171864
VERSION AR171864.1 GI:17910814
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1708)
AUTHORS Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K. and Old,L.J.
TITLE Isolated nucleic acid molecule encoding cancer associated antigen, the antigen itself, and uses thereof

```
  JOURNAL   Patent: US 6297364-A 5 02-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .1708
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 1708;  DB 6;  Length 1708;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1708;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGGGACGCTGCCGCACCGCCCCAGTTTACCCCGGGGAGCCATCATGAAGCTGAATGGCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGGACGCTGCCGCACCGCCCCAGTTTACCCCGGGGAGCCATCATGAAGCTGAATGGCCA 60

Qy         61 CCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCACAGGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCACAGGG 120

Qy        121 ACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACC 180

Qy        181 TGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGT 240

Qy        241 GCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCAC 300

Qy        301 AAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAA 360

Qy        361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGA 420

Qy        421 GATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCT 480

Qy        481 GGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGT 540

Qy        541 AGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAA 600

Qy        601 CCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGA 660

Qy        661 AATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCTCTCACCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCTCTCACCT 720

Qy        721 GATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCC 780

Qy        781 GCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGA 840

Qy        841 GCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAA 900

Qy        901 GGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACC 960

Qy        961 CGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATT 1020

Qy       1021 CAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080

Qy       1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGG 1140

Qy       1141 CAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACC 1200

Qy       1201 AAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTA 1260

Qy       1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCA 1320

Qy       1321 TCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCC 1380

Qy       1381 CTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGA 1440

Qy       1441 GAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATT 1500

Qy       1501 GAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGC 1560

Qy       1561 CCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGG 1620

Qy       1621 ATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA 1680

Qy       1681 AAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||
Db       1681 AAAATAAAATTTCCTTCAGGTTTTAAAA 1708


RESULT 2
AR343073
LOCUS       AR343073                1708 bp    DNA     linear   PAT 17-AUG-2003
DEFINITION  Sequence 5 from patent US 6576756.
ACCESSION   AR343073
VERSION     AR343073.1  GI:33738475
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1708)
  AUTHORS   Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
            and Old,L.J.
  TITLE     Isolated nucleic acid molecule encoding cancer associated antigen,
            the antigen itself, and uses thereof
  JOURNAL   Patent: US 6576756-A 5 10-JUN-2003;
FEATURES             Location/Qualifiers
     source          1. .1708
```

```
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           100.0%;  Score 1708;  DB 6;  Length 1708;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1708;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGGGACGCTGCCGCACCGCCCCAGTTTACCCCGGGGAGCCATCATGAAGCTGAATGGCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGGACGCTGCCGCACCGCCCCAGTTTACCCCGGGGAGCCATCATGAAGCTGAATGGCCA 60

Qy         61 CCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCACAGGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCACAGGG 120

Qy        121 ACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACC 180

Qy        181 TGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGT 240

Qy        241 GCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCAC 300

Qy        301 AAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAA 360

Qy        361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGA 420

Qy        421 GATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCT 480

Qy        481 GGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGT 540

Qy        541 AGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAA 600

Qy        601 CCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGA 660

Qy        661 AATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCTCTCACCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCTCTCACCT 720

Qy        721 GATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCC 780

Qy        781 GCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGA 840

Qy        841 GCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAA 900

Qy        901 GGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACC 960

Qy        961 CGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATT 1020

Qy       1021 CAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080

Qy       1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGG 1140

Qy       1141 CAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACC 1200

Qy       1201 AAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTA 1260

Qy       1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCA 1320

Qy       1321 TCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCC 1380

Qy       1381 CTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGA 1440

Qy       1441 GAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATT 1500

Qy       1501 GAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGC 1560

Qy       1561 CCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGG 1620

Qy       1621 ATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA 1680

Qy       1681 AAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||
Db       1681 AAAATAAAATTTCCTTCAGGTTTTAAAA 1708


RESULT 3
BD209924
LOCUS       BD209924               1708 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Isolated nucleic acid molecules encoding cancer-associated
            antigens, these antigens and method of using the same.
ACCESSION   BD209924
VERSION     BD209924.1  GI:33019694
KEYWORDS    JP 2002512049-A/3.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1708)
  AUTHORS   Chen,Y.T., Gure,A., Tsang,S., Stockert,E., Jager,E., Knuth,A. and
            Old,L.J.
  TITLE     Isolated nucleic acid molecules encoding cancer-associated
            antigens, these antigens and method of using the same
  JOURNAL   Patent: JP 2002512049-A 3 23-APR-2002;
            LUDWIG INSTITUTE FOR CANCER RESEARCH
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002512049-A/3
```

```
            PD   23-APR-2002
            PF   16-MAR-1999 JP 2000545030
            PR   17-APR-1998 US    09/061709
            PI   YAO TSENG CHEN,ALI GURE,SOLAM TSANG,ELISABETH STOCKERT,ELKE
            PI   JAGER,
            PI   ALEXANDER KNUTH,LLOYD J OLD
            PC   C12N15/09,A61K35/12,A61K39/00,A61K39/39,A61P35/00,C07K16/32,
            PC   C12N1/15,
            PC
            C12N1/19,C12N1/21,C12N5/10,C12P21/08,C12Q1/68,G01N33/53,G01N33/ PC
             574,
            PC   C12N15/00,C12N5/00
            CC   n=not determined(1384,1464,1533,1571,1595)
            FH   Key             Location/Qualifiers
            FT   source          1. .1708
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .1708
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           100.0%;  Score 1708;  DB 6;  Length 1708;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1708;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGGGACGCTGCCGCACCGCCCCAGTTTACCCCGGGGAGCCATCATGAAGCTGAATGGCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGGACGCTGCCGCACCGCCCCAGTTTACCCCGGGGAGCCATCATGAAGCTGAATGGCCA 60

Qy         61 CCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCACAGGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACATCCCCGATGAGCAGATAGCACAGGG 120

Qy        121 ACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCGGGGTCAGCCCCGCCAGGGCTCACC 180

Qy        181 TGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGTGGACATCCCCCTTCGGCTCCTGGT 240

Qy        241 GCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCAC 300

Qy        301 AAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAA 360

Qy        361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTGGA 420

Qy        421 GATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATCCT 480

Qy        481 GGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAGGT 540

Qy        541 AGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTACAA 600

Qy        601 CCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAGGA 660

Qy        661 AATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCTCTCACCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGCTCTCACCT 720

Qy        721 GATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGCGCAGTCCC 780

Qy        781 GCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAGGCTCCCGA 840

Qy        841 GCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATCGGCAAGAA 900

Qy        901 GGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATTGCACCACC 960

Qy        961 CGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAGGCCCAATT 1020

Qy       1021 CAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGTCCCAAGGA 1080

Qy       1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGGGTCATTGG 1140

Qy       1141 CAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTGGTAGTACC 1200

Qy       1201 AAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGACATTTCTA 1260

Qy       1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAGCAGCAGCA 1320

Qy       1321 TCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCTCCCTGTCC 1380

Qy       1381 CTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGCAGGCCTGA 1440

Qy       1441 GAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCCACTTGATT 1500

Qy       1501 GAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACCCAATTGGC 1560

Qy       1561 CCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTTAAACGTGG 1620

Qy       1621 ATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAGTGGGAAGA 1680

Qy       1681 AAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||
Db       1681 AAAATAAAATTTCCTTCAGGTTTTAAAA 1708


RESULT 4
AR171866
```

```
LOCUS       AR171866               1946 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 7 from patent US 6297364.
ACCESSION   AR171866
VERSION     AR171866.1  GI:17910816
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1946)
  AUTHORS   Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
            and Old,L.J.
  TITLE     Isolated nucleic acid molecule encoding cancer associated antigen,
            the antigen itself, and uses thereof
  JOURNAL   Patent: US 6297364-A 7 02-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .1946
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             97.9%;  Score 1672;  DB 6;  Length 1946;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1672;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 334

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 394

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 454

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 514

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 574

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        575 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 634

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        635 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 694

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        695 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 754

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        755 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 814

Qy        577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        815 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 874

Qy        637 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        875 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 934

Qy        697 TGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        935 TGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 994

Qy        757 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 816
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        995 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 1054

Qy        817 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1055 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 1114

Qy        877 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1115 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 1174

Qy        937 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 996
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1175 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 1234

Qy        997 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1056
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1235 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1294

Qy       1057 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1116
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1295 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1354

Qy       1117 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1355 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1414

Qy       1177 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1236
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1415 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1474

Qy       1237 CGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACAT 1296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1475 CGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACAT 1534

Qy       1297 CCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAG 1356
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1535 CCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAG 1594

Qy       1357 GAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGA 1416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1595 GAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGA 1654

Qy       1417 GTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTG 1476
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1655 GTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTG 1714

Qy       1477 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCA 1536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1715 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCA 1774

Qy       1537 GCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNT 1596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1775 GCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNT 1834

Qy       1597 AGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGG 1656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1835 AGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGG 1894

Qy       1657 GTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1895 GTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1946


RESULT 5
AR343075
LOCUS       AR343075               1946 bp    DNA     linear   PAT 17-AUG-2003
DEFINITION  Sequence 7 from patent US 6576756.
ACCESSION   AR343075
VERSION     AR343075.1  GI:33738477
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
```

```
            Unclassified.
REFERENCE   1  (bases 1 to 1946)
  AUTHORS   Chen,Y.-T., Gure,A., Tsang,S., Stockert,E., Jager,E., Alexander,K.
            and Old,L.J.
  TITLE     Isolated nucleic acid molecule encoding cancer associated antigen,
            the antigen itself, and uses thereof
  JOURNAL   Patent: US 6576756-A 7 10-JUN-2003;
FEATURES             Location/Qualifiers
     source          1. .1946
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             97.9%;  Score 1672;  DB 6;  Length 1946;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1672;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 334

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 394

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 454

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 514

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 574

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        575 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 634

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        635 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 694

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        695 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 754

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        755 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 814

Qy        577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        815 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 874

Qy        637 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        875 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 934

Qy        697 TGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        935 TGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 994

Qy        757 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 816
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        995 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 1054

Qy        817 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1055 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 1114

Qy        877 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1115 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 1174

Qy        937 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 996
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1175 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 1234

Qy        997 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1056
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1235 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1294

Qy       1057 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1116
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1295 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1354

Qy       1117 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1355 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1414

Qy       1177 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1236
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1415 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1474

Qy       1237 CGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACAT 1296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1475 CGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACAT 1534

Qy       1297 CCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAG 1356
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1535 CCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAG 1594

Qy       1357 GAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGA 1416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1595 GAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGA 1654

Qy       1417 GTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTG 1476
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1655 GTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTG 1714

Qy       1477 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCA 1536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1715 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCA 1774

Qy       1537 GCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNT 1596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1775 GCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNT 1834

Qy       1597 AGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGG 1656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1835 AGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGG 1894

Qy       1657 GTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1895 GTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1946

RESULT 6
BD209926
LOCUS       BD209926               1946 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Isolated nucleic acid molecules encoding cancer-associated
            antigens, these antigens and method of using the same.
ACCESSION   BD209926
VERSION     BD209926.1  GI:33019696
KEYWORDS    JP 2002512049-A/5.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1946)
  AUTHORS   Chen,Y.T., Gure,A., Tsang,S., Stockert,E., Jager,E., Knuth,A. and
            Old,L.J.
  TITLE     Isolated nucleic acid molecules encoding cancer-associated
```

```
            antigens, these antigens and method of using the same
  JOURNAL   Patent: JP 2002512049-A 5 23-APR-2002;
            LUDWIG INSTITUTE FOR CANCER RESEARCH
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002512049-A/5
            PD   23-APR-2002
            PF   16-MAR-1999 JP 2000545030
            PR   17-APR-1998 US    09/061709
            PI   YAO TSENG CHEN,ALI GURE,SOLAM TSANG,ELISABETH STOCKERT,ELKE
            PI   JAGER,
            PI   ALEXANDER KNUTH,LLOYD J OLD
            PC   C12N15/09,A61K35/12,A61K39/00,A61K39/39,A61P35/00,C07K16/32,
            PC   C12N1/15,
            PC
            C12N1/19,C12N1/21,C12N5/10,C12P21/08,C12Q1/68,G01N33/53,G01N33/ PC
             574,
            PC   C12N15/00,C12N5/00
            CC   n=not determined(1622,1702,1771,1809,1833)
            FH   Key             Location/Qualifiers
            FT   source          1. .1946
            FT                   /organism='Homo sapiens (human)'.
FEATURES              Location/Qualifiers
     source           1. .1946
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             97.9%;  Score 1672;  DB 6;  Length 1946;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1672;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 334

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 394

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 454

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 514

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        515 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 574

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        575 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 634

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        635 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 694

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        695 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 754

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        755 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 814

Qy        577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        815 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 874

Qy        637 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        875 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 934

Qy        697 TGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        935 TGTGGCTGCCATGAGCTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTT 994

Qy        757 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 816
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        995 CCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTA 1054

Qy        817 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1055 TAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGC 1114

Qy        877 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1115 AGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAG 1174

Qy        937 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 996
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1175 CGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCAT 1234

Qy        997 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1056
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1235 CACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGA 1294

Qy       1057 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1116
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1295 GGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGC 1354

Qy       1117 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1355 ATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTT 1414

Qy       1177 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1236
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1415 GACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCAT 1474

Qy       1237 CGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACAT 1296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1475 CGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACAT 1534

Qy       1297 CCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAG 1356
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1535 CCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAG 1594

Qy       1357 GAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGA 1416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1595 GAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGA 1654

Qy       1417 GTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTG 1476
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1655 GTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTG 1714

Qy       1477 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCA 1536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1715 TAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCA 1774

Qy       1537 GCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNT 1596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1775 GCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNT 1834

Qy       1597 AGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGG 1656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1835 AGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGG 1894

Qy       1657 GTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1895 GTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1946

RESULT 7
```

```
AF117106
LOCUS       AF117106               2130 bp    mRNA    linear   PRI 26-JAN-1999
DEFINITION  Homo sapiens IGF-II mRNA-binding protein 1 (IMP-1) mRNA, complete
            cds.
ACCESSION   AF117106
VERSION     AF117106.1  GI:4191607
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2130)
  AUTHORS   Nielsen,J., Christiansen,J., Lykke-Andersen,J., Johnsen,A.H.,
            Wewer,U.M. and Nielsen,F.C.
  TITLE     A family of insulin-like growth factor II mRNA-binding proteins
            represses translation in late development
  JOURNAL   Mol. Cell. Biol. 19 (2), 1262-1270 (1999)
  MEDLINE   99108099
   PUBMED   9891060
REFERENCE   2  (bases 1 to 2130)
  AUTHORS   Nielsen,J., Christiansen,J., Lykke-Andersen,J., Johnsen,A.H.,
            Wewer,U.M. and Nielsen,F.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1998) Institute of Molecular Biology, University
            of Copenhagen, Soelvgade 83H, Copenhagen DK-1307, Denmark
FEATURES             Location/Qualifiers
     source          1. .2130
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1. .2130
                     /gene="IMP-1"
     CDS             10. .1743
                     /gene="IMP-1"
                     /note="binds multiple sites in IGF-II mRNA 5'UTR"
                     /codon_start=1
                     /product="IGF-II mRNA-binding protein 1"
                     /protein_id="AAD09826.1"
                     /db_xref="GI:4191608"
                     /translation="MNKLYIGNLNESVTPADLEKVFAEHKISYSGQFLVKSGYAFVDC
                     PDEHWAMKAIETFSGKVELQGKRLEIEHSVPKKQRSRKIQIRNIPPQLRWEVLDSLLA
                     QYGTVENCEQVNTESETAVVNVTYSNREQTRQAIMKLNGHQLENHALKVSYIPDEQIA
                     QGPENGRRGGFGSRGQPRQGSPVAAGAPAKQQQVDIPLRLLVPTQYVGAIIGKEGATI
                     RNITKQTQSKIDVHRKENAGAAEKAISVHSTPEGCSSACKMILEIMHKEAKDTKTADE
                     VPLKILAHNNFVGRLIGKEGRNLKKVEQDTETKITISSLQDLTLYNPERTITVKGAIE
                     NCCRAEQEIMKKVREAYENDVAAMSLQSHLIPGLNLAAVGLFPASSSAVPPPPSSVTG
                     AAPYSSFMQAPEQEMVQVFIPAQAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSK
                     VRMVIITGPPEAQFKAQGRIYGKLKEENFFGPKEEVKLETHIRVPASAAGRVIGKGGK
                     TVNELQNLTAAEVVVPRDQTPDENDQVIVKIIGHFYASQMAQRKIRDILAQVKQQHQK
                     GQSNQAQARRK"
ORIGIN

  Query Match           58.4%;  Score 997;  DB 9;  Length 2130;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 1667;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        411 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 470

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        471 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 530

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        531 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 590

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        591 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 650

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 710

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 770

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 830

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 890

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 950

Qy        577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 1010

Qy        637 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 1070

Qy        697 TGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGG 750
              ||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db       1071 TGTGGCTGCCATGAGCCTGCAGTCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGG 1130

Qy        751 TCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGC 810
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1131 TCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGC 1190

Qy        811 TCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGC 870
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 TCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGC 1250

Qy        871 CCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTT 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1251 CCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTT 1310

Qy        931 TGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGT 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1311 TGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGT 1370

Qy        991 TATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACT 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1371 TATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACT 1430

Qy       1051 CAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGT 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1431 CAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGT 1490

Qy       1111 GCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCA 1170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1491 GCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCA 1550

Qy       1171 GAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCA 1230
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1551 GAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCA 1610

Qy       1231 GGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCG 1290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1611 GGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCG 1670

Qy       1291 AGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGC 1350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1671 AGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGC 1730

Qy       1351 ACGGAGGAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAA 1410
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
```

```
Db       1731 ACGGAGGAAGTGACCAGCCCCTCCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAA 1790

Qy       1411 TCGAGAGTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCC 1470
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1791 TCGAGAGTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCC 1850

Qy       1471 GGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGAT 1530
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1851 GGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGAT 1910

Qy       1531 CTNTCAGCCCCAAACACCCACCCAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGA 1590
             || ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1911 CTCTCAGCCCCAAACACCCACCCAATTGGCCCAACACTGTCTGCCCCTCGGGGTGTCAGA 1970

Qy       1591 AATTNTAGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACC 1650
             |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1971 AATTCTAGCGCAAGGCACTTTTAAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACC 2030

Qy       1651 AAGAGGGTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2031 AAGAGGGTGGATCACACCTCAGTGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 2088


RESULT 8
AF198254
LOCUS       AF198254               2381 bp    mRNA    linear   PRI 02-MAR-2000
DEFINITION  Homo sapiens mRNA-binding protein CRDBP mRNA, complete cds.
ACCESSION   AF198254
VERSION     AF198254.1  GI:7141071
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2381)
  AUTHORS   Ioannidis,P., Trangas,T., Dimitriadis,E., Samiotaki,M.,
            Panoutsakopoulos,G., Kyriazoglou,I., Voutzoulias,S.,
            Tsiapalis,C.M., Kittas,C., Agnantis,N. and Pandis,N.
  TITLE     Ectopic expression of a KH-domain containing protein, highly
            homologous to both human IMP-1 and mouse CRD-BP, in benign and
            malignant mesenchymal tumors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2381)
  AUTHORS   Ioannidis,P., Trangas,T., Dimitriadis,E. and Samiotaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-1999) G. Papanikolaou Research Center, St Savvas
            Hospital, 171 Alexandras Avenue, Athens, Attica 11522, Greece
FEATURES             Location/Qualifiers
     source          1. .2381
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17q21.3-q22"
                     /cell_line="MCF-7"
     CDS             306. .2039
                     /note="KH-domain containing protein; similar to Mus
                     musculus CRD-BP (c-myc mRNA coding region
                     determinant-binding protein) and Homo sapiens IMP-1
                     (insulin growth factor II mRNA-binding protein)"
                     /codon_start=1
                     /product="mRNA-binding protein CRDBP"
                     /protein_id="AAF37203.1"
                     /db_xref="GI:7141072"
                     /translation="MNKLYIGNLNESVTPADLEKVFAEHKISYSGQFLVKSGYAFVDC
                     PDEHWAMKAIETFSGKVELQGKRLEIEHSVPKKQRSRKIQIRNIPPQLRWEVLDSLLA
                     QYGTVENCEQVNTESETAVVNVTYSNREQTRQAIMKLNGHQLENHALKVSYIPDEQIA
                     QGPENGRRGGFGSRGQPRQGSPVAAGAPAKQQQVDIPLRLLVPTQYVGAIIGKEGATI
                     RNITKQTQSKIDVHRKENAGAAEKAISVHSTPEGCSSACKMILEIMHKEAKDTKTADE
                     VPLKTLAHNNFVGRLIGKEGRNLKKVEQDTETKITISSLQDLTLYNPERTITVKGAIE
                     NCCRAEQEIMKKVREAYENDVAAMSLQSHLTPGLNLAAVGLFPASSSAVPPPPSSVTG
                     AAPYSSFMQAPEQEMVQVFIPAQAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSK
                     VRMVIITGPPEAQFKAQGRIYGKLKEENFFGPKEEVKLETHIRVPASAAGRVIGKGGK
                     TVNELQNLTAAEVVVPRDQTPDENDQVIVKIIGHFYASQMAQRKIRDILAQVKQQHQK
                     GQSNQAQARRK"
ORIGIN

  Query Match           48.1%;  Score 821;  DB 9;  Length 2381;
  Best Local Similarity  99.3%;  Pred. No. 0;
  Matches 1491;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         37 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        707 AGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACAT 766

Qy         97 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 CCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTCG 826

Qy        157 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 GGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAGT 886

Qy        217 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 GGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGGA 946

Qy        277 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 GGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAA 1006

Qy        337 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 GGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTC 1066

Qy        397 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 CTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGC 1126

Qy        457 TGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 516
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1127 TGACGAGGTTCCCCTGAAGACCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAA 1186

Qy        517 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1187 GGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTC 1246

Qy        577 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 636
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1247 GTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGA 1306

Qy        637 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1307 GAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGA 1366

Qy        697 TGTGGCTGCCATGAGC------TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGG 750
              |||||||||||||||      |||||||||| ||||||||||||||||||||||||||||
Db       1367 TGTGGCTGCCATGAGCCTGCAGTCTCACCTGACCCCTGGCCTGAACCTGGCTGCTGTAGG 1426

Qy        751 TCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGC 810
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1427 TCTTTTCCCAGCTTCATCCAGCGCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGC 1486

Qy        811 TCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGC 870
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCCCTATAGCTCCTTTATGCAGGCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGC 1546

Qy        871 CCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTT 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 CCAGGCAGTGGGCGCCATCATCGGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTT 1606

Qy        931 TGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGT 990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 TGCCAGCGCCTCCATCAAGATTGCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGT 1666

Qy        991 TATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACT 1050
```

```
Db       1667 TATCATCACTGGACCGCCAGAGGCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACT 1726

Qy       1051 CAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGT 1110

Db       1727 CAAGGAGGAGAACTTCTTTGGTCCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGT 1786

Qy       1111 GCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCA 1170

Db       1787 GCCAGCATCAGCAGCTGGCCGGGTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCA 1846

Qy       1171 GAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCA 1230

Db       1847 GAATTTGACGGCAGCTGAGGTGGTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCA 1906

Qy       1231 GGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCG 1290

Db       1907 GGTCATCGTGAAAATCATCGGACATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCG 1966

Qy       1291 AGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGC 1350

Db       1967 AGACATCCTGGCCCAGGTTAAGCAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGC 2026

Qy       1351 ACGGAGGAAGTGACCAGCCCCTCCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAA 1410

Db       2027 ACGAAGGAAGTGACCAGCCCCTCCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAA 2086

Qy       1411 TCGAGAGTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCC 1470

Db       2087 TCGAGAGTGTGCTCTCCCCGGCAGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCC 2146

Qy       1471 GGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGAT 1530

Db       2147 GGGCTGTAGATCAGGTTTGCCCACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGAT 2206

Qy       1531 CT 1532

Db       2207 CT 2208
```

```
RESULT 9
BD127811
LOCUS       BD127811                2780 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD127811
VERSION     BD127811.1  GI:23222756
KEYWORDS    JP 2002017375-A/3242.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2780)
  AUTHORS   Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
            Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
            Koga,H.
  TITLE     Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002017375-A 3242 22-JAN-2002;
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002017375-A/3242
            PD   22-JAN-2002
            PF   07-JUL-2000 JP 2000253172
            PI   TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
            PI   ISHII,
            PI   YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
            SHINICHI KOJIMA,
            PI   TETSUJI OTSUKI,HISASHI KOGA
            PC
            C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
            PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
             Primer for synthesizing full-length cDNA and use thereof FH   Key
                      Location/Qualifiers
```

```
            FT   CDS             (112). .(1068).
FEATURES             Location/Qualifiers
     source          1. .2780
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           43.4%;  Score 741;  DB 6;  Length 2780;
  Best Local Similarity  99.5%;  Pred. No. 0;
  Matches  991;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772

Db        421 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 480

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832

Db        481 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 540

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892

Db        541 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 600

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952

Db        601 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 660

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012

Db        661 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 720

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072

Db        721 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 780

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132

Db        781 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 840

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192

Db        841 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 900

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252

Db        901 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 960

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312

Db        961 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1020

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1372

Db       1021 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1080

Qy       1373 CCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1432

Db       1081 CCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1140

Qy       1433 AGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCC 1492

Db       1141 AGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCC 1200

Qy       1493 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACC 1552

Db       1201 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACC 1260

Qy       1553 CAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTT 1612

Db       1261 CAATTGGCCCAACACTGTCTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTT 1320

Qy       1613 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1672
```

```
Db       1321 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1380

Qy       1673 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
Db       1381 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1416
```

RESULT 10
AK074915
LOCUS AK074915 2780 bp mRNA linear PRI 03-SEP-2002
DEFINITION Homo sapiens cDNA FLJ90434 fis, clone NT2RP3000789, highly similar to Homo sapiens IGF-II mRNA-binding protein 1 (IMP-1) mRNA.
ACCESSION AK074915
VERSION AK074915.1 GI:22760672
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T., Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S., Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Kojima,S., Nagahari,K., Masuho,Y., Ono,T., Okano,K., Yoshikawa,Y., Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and Ninomiya,K.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2780)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).
FEATURES Location/Qualifiers
source 1. .2780
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP3000789"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP3"
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"
ORIGIN

Query Match 43.4%; Score 741; DB 9; Length 2780;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 991; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
Db        421 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 480

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
Db        481 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 540

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
Db        541 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 600

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
Db        601 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 660

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
Db        661 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 720

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
Db        721 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 780

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
Db        781 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 840

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
Db        841 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 900

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
Db        901 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 960

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
Db        961 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1020

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1372
Db       1021 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAGTGACCAGCCCCT 1080

Qy       1373 CCCTGTCCCTTNGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1432
Db       1081 CCCTGTCCCTTCGAGTCCAGGACAACAACGGGCAGAAATCGAGAGTGTGCTCTCCCCGGC 1140

Qy       1433 AGGCCTGAGAATGAGTGGGAATCCGGGACACNTGGGCCGGGCTGTAGATCAGGTTTGCCC 1492
Db       1141 AGGCCTGAGAATGAGTGGGAATCCGGGACACCTGGGCCGGGCTGTAGATCAGGTTTGCCC 1200

Qy       1493 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTNTCAGCCCCAAACACCCACC 1552
Db       1201 ACTTGATTGAGAAAGATGTTCCAGTGAGGAACCCTGATCTCTCAGCCCCAAACACCCACC 1260

Qy       1553 CAATTGGCCCAACACTGTNTGCCCCTCGGGGTGTCAGAAATTNTAGCGCAAGGCACTTTT 1612
Db       1261 CAATTGGCCCAACACTGTCTGCCCCTCGGGGTGTCAGAAATTCTAGCGCAAGGCACTTTT 1320

Qy       1613 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1672
Db       1321 AAACGTGGATTGTTTAAAGAAGCTCTCCAGGCCCCACCAAGAGGGTGGATCACACCTCAG 1380

Qy       1673 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1708
Db       1381 TGGGAAGAAAAATAAAATTTCCTTCAGGTTTTAAAA 1416
```

RESULT 11
BD124998
LOCUS BD124998 833 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD124998
VERSION BD124998.1 GI:23219943
KEYWORDS JP 2002017375-A/429.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 833)
AUTHORS Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017375-A 429 22-JAN-2002;

```
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002017375-A/429
            PD   22-JAN-2002
            PF   07-JUL-2000 JP 2000253172
            PI   TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
            PI   ISHII,
            PI   YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
            SHINICHI KOJIMA,
            PI   TETSUJI OTSUKI,HISASHI KOGA
            PC
            C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
            PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
             Primer for synthesizing full-length cDNA and use thereof FH   Key
                     Location/Qualifiers
            FT   source          1. .833
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .833
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             24.2%;  Score 414;  DB 6;  Length 833;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-213;
  Matches  414;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        299 ACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAA 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAA 60

Qy        359 AAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTG 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTG 120

Qy        419 GAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATC 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATC 180

Qy        479 CTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAG 538
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAG 240

Qy        539 GTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTAC 598
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTAC 300

Qy        599 AACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAG 658
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAG 360

Qy        659 GAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 414


RESULT 12
BD126684
LOCUS       BD126684                 833 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD126684
VERSION     BD126684.1  GI:23221629
KEYWORDS    JP 2002017375-A/2115.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 833)
  AUTHORS   Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
            Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
            Koga,H.
  TITLE     Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002017375-A 2115 22-JAN-2002;
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002017375-A/2115
            PD   22-JAN-2002
            PF   07-JUL-2000 JP 2000253172
            PI   TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
            PI   ISHII,
            PI   YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
            SHINICHI KOJIMA,
            PI   TETSUJI OTSUKI,HISASHI KOGA
            PC
            C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
            PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
             Primer for synthesizing full-length cDNA and use thereof FH   Key
                     Location/Qualifiers
            FT   source          1. .833
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .833
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             24.2%;  Score 414;  DB 6;  Length 833;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-213;
  Matches  414;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        299 ACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAA 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACAAAACAGACCCAGTCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAA 60

Qy        359 AAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTG 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGCTCCTCCGCTTGTAAGATGATCTTG 120

Qy        419 GAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATC 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGATTATGCATAAAGAGGCTAAGGACACCAAAACGGCTGACGAGGTTCCCCTGAAGATC 180

Qy        479 CTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAG 538
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGCAAGGAAGGACGGAACCTGAAGAAG 240

Qy        539 GTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTAC 598
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTAGAGCAAGATACCGAGACAAAAATCACCATCTCCTCGTTGCAAGACCTTACCCTTTAC 300

Qy        599 AACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAG 658
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AACCCTGAGAGGACCATCACTGTGAAGGGGGCCATCGAGAATTGTTGCAGGGCCGAGCAG 360

Qy        659 GAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAATGATGTGGCTGCCATGAGC 414


RESULT 13
AC105030/c
LOCUS       AC105030               107848 bp    DNA     linear   PRI 05-NOV-2002
DEFINITION  Homo sapiens chromosome 17, clone CTD-2244F11, complete sequence.
ACCESSION   AC105030
VERSION     AC105030.11  GI:24580458
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 107848)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
```

```
  TITLE     Homo sapiens chromosome 17, clone CTD-2244F11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 107848)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 107848)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 107848)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 5, 2002 this sequence version replaced gi:23506732.

            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23011
                Center clone name: 2244_F_11
            --------------

            Only the initial 107.85 kb of this clone are being submitted.
            The remainder of the clone is overlapped by accession number
            AC091133 [WICGR project L12028].
FEATURES                Location/Qualifiers
     source             1. .107848
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="17"
                        /map="17"
                        /clone="CTD-2244F11"
                        /clone_lib="CITD1 Human BAC"
     repeat_region      1. .160
                        /rpt_family="LTR19A"
     repeat_region      784. .860
                        /rpt_family="L2"
     repeat_region      1215. .1493
                        /rpt_family="AluSx"
     repeat_region      1498. .1559
                        /rpt_family="(CA)n"
     repeat_region      1591. .1887
                        /rpt_family="L2"
     repeat_region      2336. .2530
                        /rpt_family="MIR"
     repeat_region      2635. .2937
                        /rpt_family="AluSx"
     repeat_region      3488. .3516
                        /rpt_family="AT_rich"
     repeat_region      3517. .3732
                        /rpt_family="AluJb"
     repeat_region      complement(3798. .4269)
                        /rpt_family="LTR33"
     repeat_region      complement(4320. .4414)
                        /rpt_family="MLT1J1"
     repeat_region      4448. .4744
                        /rpt_family="AluSx"
     repeat_region      complement(4745. .4791)
                        /rpt_family="MLT1J2"
     repeat_region      complement(4868. .4961)
                        /rpt_family="MLT1J1"
     repeat_region      5740. .5914
                        /rpt_family="LTR44"
     repeat_region      6101. .6244
                        /rpt_family="AluJo"
     repeat_region      7572. .7962
                        /rpt_family="MER57-int"
     repeat_region      complement(8024. .8187)
                        /rpt_family="L1MA4"
     repeat_region      8188. .8484
                        /rpt_family="AluSg"
     repeat_region      complement(8485. .8605)
                        /rpt_family="L1MA4"
     repeat_region      8607. .8878
                        /rpt_family="L1MA4"
     repeat_region      9844. .9992
                        /rpt_family="MER57-int"
     repeat_region      10087. .10397
                        /rpt_family="AluJb"
     repeat_region      complement(10417. .10647)
```

```
                     /rpt_family="MER30"
     repeat_region   11922. .12037
                     /rpt_family="FLAM_C"
     repeat_region   complement(12063. .12263)
                     /rpt_family="AluJo"
     repeat_region   13037. .13140
                     /rpt_family="MER57-int"
     repeat_region   complement(13679. .13786)
                     /rpt_family="AluSg/x"
     repeat_region   13816. .14271
                     /rpt_family="LTR44"
     repeat_region   14272. .14393
                     /rpt_family="FLAM_A"
     repeat_region   14394. .14434
                     /rpt_family="(TG)n"
     repeat_region   14442. .14477
                     /rpt_family="LTR44"
     repeat_region   14566. .14803
                     /rpt_family="MIR"
     repeat_region   15796. .15877
                     /rpt_family="MIR"
     repeat_region   15989. .16046
                     /rpt_family="L3"
     repeat_region   16215. .16602
                     /rpt_family="L2"
     repeat_region   complement(16724. .17021)
                     /rpt_family="AluSx"
     repeat_region   17048. .17408
                     /rpt_family="MLT1H"
     repeat_region   17383. .17414
                     /rpt_family="MLT1H"
     repeat_region   complement(17510. .17815)
                     /rpt_family="AluJb"
     repeat_region   18437. .18716
                     /rpt_family="AluSc"
     repeat_region   18731. .19060
                     /rpt_family="(TA)n"
     repeat_region   19100. .19261
                     /rpt_family="(TA)n"
     repeat_region   19299. .19323
                     /rpt_family="(CA)n"
     repeat_region   complement(19513. .19698)
                     /rpt_family="AluJb"

  Query Match             16.6%;  Score 284;  DB 9;  Length 107848;
  Best Local Similarity   100.0%;  Pred. No. 1e-142;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         36 GAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACA 95
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83857 GAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACA 83798

Qy         96 TCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTC 155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83797 TCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTC 83738

Qy        156 GGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAG 215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83737 GGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAG 83678

Qy        216 TGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGG 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83677 TGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGG 83618

Qy        276 AGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAG 319
              |||||||||||||||||||||||||||||||||||||||||||
Db      83617 AGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAG 83574


RESULT 14
AC104974
LOCUS       AC104974              154604 bp    DNA     linear   HTG 13-FEB-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-145L16 map 17, WORKING DRAFT
            SEQUENCE, 23 ordered pieces.
ACCESSION   AC104974
VERSION     AC104974.3  GI:18653683
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154604)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-145L16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154604)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 13, 2002 this sequence version replaced gi:18464184.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22871
                Center clone name: 145_L_16
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 148031 bases at least Q40
                Consensus quality: 150405 bases at least Q30
                Consensus quality: 151462 bases at least Q20
                Insert size: 146000; agarose-fp
                Insert size: 152404; sum-of-contigs
                Quality coverage: 9.3 in Q20 bases; agarose-fp
                Quality coverage: 8.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1     1709: contig of 1709 bp in length
            *    1710     1809: gap of 100 bp
```

```
            *     1810     1999: contig of 190 bp in length
            *     2000     2099: gap of 100 bp
            *     2100     3127: contig of 1028 bp in length
            *     3128     3227: gap of 100 bp
            *     3228     4561: contig of 1334 bp in length
            *     4562     4661: gap of 100 bp
            *     4662     5811: contig of 1150 bp in length
            *     5812     5911: gap of 100 bp
            *     5912     7151: contig of 1240 bp in length
            *     7152     7251: gap of 100 bp
            *     7252     8261: contig of 1010 bp in length
            *     8262     8361: gap of 100 bp
            *     8362     9739: contig of 1378 bp in length
            *     9740     9839: gap of 100 bp
            *     9840    12105: contig of 2266 bp in length
            *    12106    12205: gap of 100 bp
            *    12206    35760: contig of 23555 bp in length
            *    35761    35860: gap of 100 bp
            *    35861    38000: contig of 2140 bp in length
            *    38001    38100: gap of 100 bp
            *    38101    41369: contig of 3269 bp in length
            *    41370    41469: gap of 100 bp
            *    41470    44094: contig of 2625 bp in length
            *    44095    44194: gap of 100 bp
            *    44195    47400: contig of 3206 bp in length
            *    47401    47500: gap of 100 bp
            *    47501    53741: contig of 6241 bp in length
            *    53742    53841: gap of 100 bp
            *    53842    60833: contig of 6992 bp in length
            *    60834    60933: gap of 100 bp
            *    60934    72421: contig of 11488 bp in length
            *    72422    72521: gap of 100 bp
            *    72522    84887: contig of 12366 bp in length
            *    84888    84987: gap of 100 bp
            *    84988    99600: contig of 14613 bp in length
            *    99601    99700: gap of 100 bp
            *    99701   115353: contig of 15653 bp in length
            *   115354   115453: gap of 100 bp
            *   115454   131255: contig of 15802 bp in length
            *   131256   131355: gap of 100 bp
            *   131356   152505: contig of 21150 bp in length
            *   152506   152605: gap of 100 bp
            *   152606   154604: contig of 1999 bp in length.
FEATURES             Location/Qualifiers
     source          1. .154604
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-145L16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1709
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1810. .1999
                     /note="assembly_fragment"
     misc_feature    2100. .3127
                     /note="assembly_fragment"
     misc_feature    3228. .4561
                     /note="assembly_fragment"
     misc_feature    4662. .5811
                     /note="assembly_fragment"
     misc_feature    5912. .7151
                     /note="assembly_fragment"
     misc_feature    7252. .8261
                     /note="assembly_fragment"
     misc_feature    8362. .9739
                     /note="assembly_fragment"
     misc_feature    9840. .12105
                     /note="assembly_fragment"
     misc_feature    12206. .35760
                     /note="assembly_fragment"
     misc_feature    35861. .38000
                     /note="assembly_fragment"
     misc_feature    38101. .41369
                     /note="assembly_fragment"
     misc_feature    41470. .44094
                     /note="assembly_fragment"
     misc_feature    44195. .47400
                     /note="assembly_fragment"
     misc_feature    47501. .53741
                     /note="assembly_fragment"
     misc_feature    53842. .60833
                     /note="assembly_fragment"
     misc_feature    60934. .72421
                     /note="assembly_fragment"
     misc_feature    72522. .84887
                     /note="assembly_fragment"
     misc_feature    84988. .99600
                     /note="assembly_fragment"
     misc_feature    99701. .115353
                     /note="assembly_fragment"
     misc_feature    115454. .131255
                     /note="assembly_fragment"
     misc_feature    131356. .152505
                     /note="assembly_fragment"
     misc_feature    152606. .154604
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             16.6%;  Score 284;  DB 2;  Length 154604;
  Best Local Similarity   100.0%;  Pred. No. 1e-142;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         36 GAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACA 95
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133545 GAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACA 133604

Qy         96 TCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTC 155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133605 TCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTC 133664

Qy        156 GGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAG 215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133665 GGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAG 133724

Qy        216 TGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGG 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133725 TGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGG 133784

Qy        276 AGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAG 319
              |||||||||||||||||||||||||||||||||||||||||||
Db     133785 AGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAG 133828


RESULT 15
AC025556/c
LOCUS       AC025556              159122 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-501C14, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC025556
VERSION     AC025556.4  GI:9958273
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159122)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159122)
```

```
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:8570253.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0501C14
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 96%
            Sequencing vector: plasmid; 4%
            Chemistry: Dye-primer ET; 96% of reads
            Chemistry: Dye-terminator Big Dye; 4% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 148686 bases at least Q40
            Consensus quality: 152251 bases at least Q30
            Consensus quality: 154277 bases at least Q20
            Insert size: 181000; agarose-fp
            Insert size: 157522; sum-of-contigs
            Quality coverage: 3.86 in Q20 bases; agarose-fp
            Quality coverage: 4.48 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7635: contig of 7635 bp in length
            *     7636     7735: gap of unknown length
            *     7736    15480: contig of 7745 bp in length
            *    15481    15580: gap of unknown length
            *    15581    25089: contig of 9509 bp in length
            *    25090    25189: gap of unknown length
            *    25190    35006: contig of 9817 bp in length
            *    35007    35106: gap of unknown length
            *    35107    44509: contig of 9403 bp in length
            *    44510    44609: gap of unknown length
            *    44610    56601: contig of 11992 bp in length
            *    56602    56701: gap of unknown length
            *    56702    64536: contig of 7835 bp in length
            *    64537    64636: gap of unknown length
            *    64637    79941: contig of 15305 bp in length
            *    79942    80041: gap of unknown length
            *    80042    95746: contig of 15705 bp in length
            *    95747    95846: gap of unknown length
            *    95847   125532: contig of 29686 bp in length
            *   125533   125632: gap of unknown length
            *   125633   126811: contig of 1179 bp in length
            *   126812   126911: gap of unknown length
            *   126912   129146: contig of 2235 bp in length
            *   129147   129246: gap of unknown length
            *   129247   132692: contig of 3446 bp in length
            *   132693   132792: gap of unknown length
            *   132793   135914: contig of 3122 bp in length
            *   135915   136014: gap of unknown length
            *   136015   143560: contig of 7546 bp in length
            *   143561   143660: gap of unknown length
            *   143661   150766: contig of 7106 bp in length
            *   150767   150866: gap of unknown length
            *   150867   159122: contig of 8256 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159122
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-501C14"
     misc_feature    1. .7635
                     /note="assembly_name:Contig10"
     misc_feature    7736. .15480
                     /note="assembly_name:Contig11"
     misc_feature    15581. .25089
                     /note="assembly_name:Contig12"
     misc_feature    25190. .35006
                     /note="assembly_name:Contig13"
     misc_feature    35107. .44509
                     /note="assembly_name:Contig14"
     misc_feature    44610. .56601
                     /note="assembly_name:Contig15"
     misc_feature    56702. .64536
                     /note="assembly_name:Contig16"
     misc_feature    64637. .79941
                     /note="assembly_name:Contig17"
     misc_feature    80042. .95746
                     /note="assembly_name:Contig18"
     misc_feature    95847. .125532
                     /note="assembly_name:Contig19"
     misc_feature    125633. .126811
                     /note="assembly_name:Contig3"
     misc_feature    126912. .129146
                     /note="assembly_name:Contig4"
     misc_feature    129247. .132692
                     /note="assembly_name:Contig6"
     misc_feature    132793. .135914
                     /note="assembly_name:Contig5"
     misc_feature    136015. .143560
                     /note="assembly_name:Contig7"
     misc_feature    143661. .150766
                     /note="assembly_name:Contig8"
     misc_feature    150867. .159122
                     /note="assembly_name:Contig9"
ORIGIN

  Query Match             16.6%;  Score 284;  DB 2;  Length 159122;
  Best Local Similarity   100.0%;  Pred. No. 1e-142;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          36 GAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACA 95
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       67231 GAGCCATCATGAAGCTGAATGGCCACCAGTTGGAGAACCATGCCCTGAAGGTCTCCTACA 67172

Qy          96 TCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTC 155
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       67171 TCCCCGATGAGCAGATAGCACAGGGACCTGAGAATGGGCGCCGAGGGGGCTTTGGCTCTC 67112

Qy         156 GGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAG 215
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       67111 GGGGTCAGCCCCGCCAGGGCTCACCTGTGGCAGCGGGGGCCCCAGCCAAGCAGCAGCAAG 67052

Qy         216 TGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGG 275
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       67051 TGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAGG 66992

Qy         276 AGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAG 319
               |||||||||||||||||||||||||||||||||||||||||||
Db       66991 AGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAG 66948


Search completed: July 14, 2004, 01:45:04
Job time : 6733 secs
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 13, 2004, 12:13:53 ; Search time 81.5 Seconds
(without alignments)
11842.727 Million cell updates/sec

Title: US-09-270-437D-5
Perfect score: 560
Sequence: 1 agggacgctgccgcaccgcc.....atttccttcaggttttaaaa 1708

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2987630

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09270437/runat_13072004_121955_9961/app_query.fasta_1.1863
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09270437_@CGN_1_1_154_@runat_13072004_121955_9961 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	216	38.6	261	4	AAU16161	Aau16161 Human nov
2	216	38.6	261	6	ABU55230	Abu55230 Human nov
3	216	38.6	319	4	AAM93826	Aam93826 Human pol
4	166	29.6	577	2	AAY30649	Aay30649 A murine
5	116	20.7	250	4	AAU16579	Aau16579 Human nov
6	116	20.7	250	6	ABU55648	Abu55648 Human nov
7	36	6.4	579	3	AAB11328	Aab11328 Human lun
8	36	6.4	579	3	AAB11365	Aab11365 Human lun
9	36	6.4	579	5	ABB74960	Abb74960 Human lun
10	36	6.4	579	5	ABB75053	Abb75053 Human lun
11	36	6.4	579	5	ABB74997	Abb74997 Human lun
12	36	6.4	579	5	ABB75054	Abb75054 Human lun
13	36	6.4	579	5	ABP61917	Abp61917 Human lun
14	36	6.4	579	5	ABP61974	Abp61974 Human lun
15	36	6.4	579	5	ABP61880	Abp61880 Human lun
16	36	6.4	579	5	ABP61973	Abp61973 Human lun
17	36	6.4	579	7	ADA28536	Ada28536 Recombina
18	36	6.4	579	7	ADA28539	Ada28539 Recombina
19	36	6.4	579	7	ADA28438	Ada28438 Human lun
20	36	6.4	579	7	ADA28266	Ada28266 Human lun
21	36	6.4	579	7	ADD14066	Add14066 Human src
22	36	6.4	579	7	ADE53471	Ade53471 Human lun
23	36	6.4	583	4	ABG12592	Abg12592 Novel hum
24	36	6.4	586	5	ABB75048	Abb75048 Human lun
25	36	6.4	586	5	ABP61968	Abp61968 Human lun
26	36	6.4	586	7	ADA28517	Ada28517 Recombina
27	36	6.4	619	4	ABG21963	Abg21963 Novel hum
28	24	4.3	148	4	ABG21962	Abg21962 Novel hum
29	21	3.8	93	4	AAM38501	Aam38501 Peptide #
30	21	3.8	93	4	AAM78238	Aam78238 Human bon
31	20	3.6	20	5	ABB75042	Abb75042 Human lun
32	20	3.6	20	5	ABB75041	Abb75041 Human lun
33	20	3.6	20	5	ABP61961	Abp61961 Human lun
34	20	3.6	20	5	ABP61962	Abp61962 Human lun
35	20	3.6	20	7	ADA28504	Ada28504 Human lun
36	20	3.6	20	7	ADA28505	Ada28505 Human lun
37	19	3.4	171	4	AAU16166	Aau16166 Human nov
38	19	3.4	171	4	AAU16583	Aau16583 Human nov
39	19	3.4	171	6	ABU55235	Abu55235 Human nov
40	19	3.4	171	6	ABU55652	Abu55652 Human nov
41	19	3.4	556	5	ABG96346	Abg96346 Human ova
42	19	3.4	594	4	ABG06795	Abg06795 Novel hum
43	19	3.4	614	4	ABG06794	Abg06794 Novel hum
44	19	3.4	620	4	AAU16163	Aau16163 Human nov
45	19	3.4	620	6	ABU55232	Abu55232 Human nov

ALIGNMENTS

RESULT 1
AAU16161
ID AAU16161 standard; protein; 261 AA.
XX
AC AAU16161;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 1114.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS26148.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
PS Claim 11; SEQ ID NO 1114; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present sequence represents a novel secreted protein of the invention. Note: The sequence data for this patent did not form part of the printed

Alignment Scores:

Pred. No.:	6.67e-204	Length:	261
Score:	216.00	Matches:	216
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	38.57%	Indels:	0
DB:	4	Gaps:	0

US-09-270-437D-5 (1-1708) x AAU16161 (1-261)

```
Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 SerHisLeuIleProGlyLeuAsnLeuAlaAlaValGlyLeuPheProAlaSerSerSer 65

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 AlaValProProProProSerSerValThrGlyAlaAlaProTyrSerSerPheMetGln 85

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 AlaProGluGlnGluMetValGlnValPheIleProAlaGlnAlaValGlyAlaIleIle 105

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 GlyLysLysGlyGlnHisIleLysGlnLeuSerArgPheAlaSerAlaSerIleLysIle 125

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 AlaProProGluThrProAspSerLysValArgMetValIleIleThrGlyProProGlu 145

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 AlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGluGluAsnPhePheGly 165

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 ProLysGluGluValLysLeuGluThrHisIleArgValProAlaSerAlaAlaGlyArg 185

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 ValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrAlaAlaGluVal 205

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 ValValProArgAspGlnThrProAspGluAsnAspGlnValIleValLysIleIleGly 225

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 HisPheTyrAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAlaGlnValLys 245

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAG 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GlnGlnHisGlnLysGlyGlnSerAsnGlnAlaGlnAlaArgArgLys 261
```

RESULT 2
ABU55230
ID ABU55230 standard; protein; 261 AA.
XX
AC ABU55230;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polypeptide #317.
XX
KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-00764864.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.

```
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73489.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 11; SEQ ID NO 1114; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC   ABU55748 represent human novel polypeptides of the invention
XX
SQ   Sequence 261 AA;

Alignment Scores:
Pred. No.:              6.67e-204      Length:         261
Score:                  216.00         Matches:        216
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            38.57%         Indels:         0
DB:                     6              Gaps:           0

US-09-270-437D-5 (1-1708) x ABU55230 (1-261)

Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 SerHisLeuIleProGlyLeuAsnLeuAlaAlaValGlyLeuProAlaSerSerSer 65

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 AlaValProProProProSerSerValThrGlyAlaAlaProTyrSerSerPheMetGln 85

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 AlaProGluGlnGluMetValGlnValPheIleProAlaGlnAlaValGlyAlaIleIle 105

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 GlyLysLysGlyGlnHisIleLysGlnLeuSerArgPheAlaSerAlaSerIleLysIle 125

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 AlaProProGluThrProAspSerLysValArgMetValIleIleThrGlyProProGlu 145

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 AlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGluGluAsnPhePheGly 165

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 ProLysGluGluValLysLeuGluThrHisIleArgValProAlaSerAlaAlaGlyArg 185

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 ValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrAlaAlaGluVal 205

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 ValValProArgAspGlnThrProAspGluAsnAspGlnValIleValLysIleIleGly 225

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 HisPheTyrAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAlaGlnValLys 245

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAG 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GlnGlnHisGlnLysGlyGlnSerAsnGlnAlaGlnAlaArgArgLys 261

RESULT 3
AAM93826
ID   AAM93826 standard; protein; 319 AA.
XX
AC   AAM93826;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide, SEQ ID NO: 3887.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-00114089.
XX
PR   08-JUL-1999;   99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
DR   N-PSDB; AAK94782.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their use
PT   in genetic manipulation.
XX
PS   Claim 8; SEQ ID NO 3887; 1380pp + Sequence Listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been isolated
CC   and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC   been determined. Primers for synthesising the full length cDNA are useful
CC   for clarifying the function of the protein encoded by the cDNA. The full
CC   length clones were obtained by construction of full length enriched cDNA
CC   libraries that were synthesised by the oligo-capping method. The primers
CC   enable the production of the full length cDNA easily without any special
CC   methods. The present sequence is a polypeptide encoded by a full length
CC   human cDNA of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in CD-ROM
CC   format directly from EPO
XX
```

SQ Sequence 319 AA;

Alignment Scores:

Pred. No.:	6.46e-204	Length:	319
Score:	216.00	Matches:	216
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	38.57%	Indels:	0
DB:	4	Gaps:	0

US-09-270-437D-5 (1-1708) x AAM93826 (1-319)

```
Qy        713 TCTCACCTGATCCCTGGCCTGAACCTGGCTGCTGTAGGTCTTTTCCCAGCTTCATCCAGC 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 SerHisLeuIleProGlyLeuAsnLeuAlaAlaValGlyLeuPheProAlaSerSerSer 123

Qy        773 GCAGTCCCGCCGCCTCCCAGCAGCGTTACTGGGGCTGCTCCCTATAGCTCCTTTATGCAG 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 AlaValProProProProSerSerValThrGlyAlaAlaProTyrSerSerPheMetGln 143

Qy        833 GCTCCCGAGCAGGAGATGGTGCAGGTGTTTATCCCCGCCCAGGCAGTGGGCGCCATCATC 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 AlaProGluGlnGluMetValGlnValPheIleProAlaGlnAlaValGlyAlaIleIle 163

Qy        893 GGCAAGAAGGGGCAGCACATCAAACAGCTCTCCCGGTTTGCCAGCGCCTCCATCAAGATT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 GlyLysLysGlyGlnHisIleLysGlnLeuSerArgPheAlaSerAlaSerIleLysIle 183

Qy        953 GCACCACCCGAAACACCTGACTCCAAAGTTCGTATGGTTATCATCACTGGACCGCCAGAG 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 AlaProProGluThrProAspSerLysValArgMetValIleIleThrGlyProProGlu 203

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 AlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGluGluAsnPhePheGly 223

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 ProLysGluGluValLysLeuGluThrHisIleArgValProAlaSerAlaAlaGlyArg 243

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 ValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrAlaAlaGluVal 263

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 ValValProArgAspGlnThrProAspGluAsnAspGlnValIleValLysIleIleGly 283

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        284 HisPheTyrAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAlaGlnValLys 303

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAG 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 GlnGlnHisGlnLysGlyGlnSerAsnGlnAlaGlnAlaArgArgLys 319
```

RESULT 4
AAY30649
ID AAY30649 standard; protein; 577 AA.
XX
AC AAY30649;
XX
DT 17-NOV-1999 (first entry)
XX
DE A murine c-myc coding region determinant binding protein.
XX
KW c-myc coding region determinant binding protein; CRD-BP; tumor; c-myc;
KW endonucleolytic attack; half-life; breast cancer; colon cancer;
KW pancreatic cancer.
XX
OS Mus musculus.
XX
PN WO9946594-A2.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004897.
XX
PR 09-MAR-1998; 98US-0077372P.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Ross J;
XX
DR WPI; 1999-551506/46.
DR N-PSDB; AAZ10617.
XX
PT Diagnosing presence or absence of a tumor in a human by examining c-myc
PT coding region determinant-binding protein.
XX
PS Example; Fig 1A-D; 79pp; English.
XX
CC The present sequence represents a murine c-myc coding region determinant
CC binding protein (CRD-BP). The presence or absence of a tumor can be
CC determined by determining the levels of CRD-BP present in the suspect
CC tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack
CC and so prolongs its half-life. The methods are used for diagnosing
CC presence or absence of a tumor in a human, especially breast, colon and
CC pancreatic cancer. They are also used to inhibit cancer cell growth
XX
SQ Sequence 577 AA;

Alignment Scores:

Pred. No.:	1.45e-154	Length:	577
Score:	166.00	Matches:	166
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	29.64%	Indels:	0
DB:	2	Gaps:	0

US-09-270-437D-5 (1-1708) x AAY30649 (1-577)

```
Qy        215 GTGGACATCCCCCTTCGGCTCCTGGTGCCCACCCAGTATGTGGGTGCCATTATTGGCAAG 274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 ValAspIleProLeuArgLeuLeuValProThrGlnTyrValGlyAlaIleIleGlyLys 213

Qy        275 GAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAGTCCAAGATAGACGTGCATAGG 334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 GluGlyAlaThrIleArgAsnIleThrLysGlnThrGlnSerLysIleAspValHisArg 233

Qy        335 AAGGAGAACGCAGGTGCAGCTGAAAAAGCCATCAGTGTGCACTCCACCCCTGAGGGCTGC 394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        234 LysGluAsnAlaGlyAlaAlaGluLysAlaIleSerValHisSerThrProGluGlyCys 253

Qy        395 TCCTCCGCTTGTAAGATGATCTTGGAGATTATGCATAAAGAGGCTAAGGACACCAAAACG 454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 SerSerAlaCysLysMetIleLeuGluIleMetHisLysGluAlaLysAspThrLysThr 273

Qy        455 GCTGACGAGGTTCCCCTGAAGATCCTGGCCCATAATAACTTTGTAGGGCGTCTCATTGGC 514
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 AlaAspGluValProLeuLysIleLeuAlaHisAsnAsnPheValGlyArgLeuIleGly 293

Qy        515 AAGGAAGGACGGAACCTGAAGAAGGTAGAGCAAGATACCGAGACAAAAATCACCATCTCC 574
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 LysGluGlyArgAsnLeuLysLysValGluGlnAspThrGluThrLysIleThrIleSer 313

Qy        575 TCGTTGCAAGACCTTACCCTTTACAACCCTGAGAGGACCATCACTGTGAAGGGGGCCATC 634
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 SerLeuGlnAspLeuThrLeuTyrAsnProGluArgThrIleThrValLysGlyAlaIle 333

Qy        635 GAGAATTGTTGCAGGGCCGAGCAGGAAATAATGAAGAAAGTTCGGGAGGCCTATGAGAAT 694
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 GluAsnCysCysArgAlaGluGlnGluIleMetLysLysValArgGluAlaTyrGluAsn 353

Qy        695 GATGTGGCTGCCATGAGC 712
              ||||||||||||||||||
```

Db 354 AspValAlaAlaMetSer 359

RESULT 5
AAU16579
ID AAU16579 standard; protein; 250 AA.
XX
AC AAU16579;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 1532.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.

PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS26566.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
PS Claim 11; SEQ ID NO 1532; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Alignment Scores:

Pred. No.:	4.12e-105	Length:	250
Score:	116.00	Matches:	116
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	20.71%	Indels:	0
DB:	4	Gaps:	0

US-09-270-437D-5 (1-1708) x AAU16579 (1-250)

```
Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 AlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGluGluAsnPhePheGly 154

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 ProLysGluGluValLysLeuGluThrHisIleArgValProAlaSerAlaAlaGlyArg 174

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        175 ValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrAlaAlaGluVal 194

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 ValValProArgAspGlnThrProAspGluAsnAspGlnValIleValLysIleIleGly 214

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 HisPheTyrAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAlaGlnValLys 234

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAG 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 GlnGlnHisGlnLysGlyGlnSerAsnGlnAlaGlnAlaArgArgLys 250
```

RESULT 6
ABU55648
ID ABU55648 standard; protein; 250 AA.
XX
AC ABU55648;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polypeptide #735.
XX
KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-00764864.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.

```
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73907.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 11; SEQ ID NO 1532; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC   ABU55748 represent human novel polypeptides of the invention
XX
SQ   Sequence 250 AA;

Alignment Scores:
Pred. No.:                4.12e-105      Length:          250
Score:                    116.00         Matches:         116
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              20.71%         Indels:          0
DB:                       6              Gaps:            0

US-09-270-437D-5 (1-1708) x ABU55648 (1-250)

Qy       1013 GCCCAATTCAAGGCTCAGGGAAGAATCTATGGCAAACTCAAGGAGGAGAACTTCTTTGGT 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 AlaGlnPheLysAlaGlnGlyArgIleTyrGlyLysLeuLysGluGluAsnPhePheGly 154

Qy       1073 CCCAAGGAGGAAGTGAAGCTGGAGACCCACATACGTGTGCCAGCATCAGCAGCTGGCCGG 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 ProLysGluGluValLysLeuGluThrHisIleArgValProAlaSerAlaAlaGlyArg 174

Qy       1133 GTCATTGGCAAAGGTGGAAAAACGGTGAACGAGTTGCAGAATTTGACGGCAGCTGAGGTG 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        175 ValIleGlyLysGlyGlyLysThrValAsnGluLeuGlnAsnLeuThrAlaAlaGluVal 194

Qy       1193 GTAGTACCAAGAGACCAGACCCCTGATGAGAACGACCAGGTCATCGTGAAAATCATCGGA 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 ValValProArgAspGlnThrProAspGluAsnAspGlnValIleValLysIleIleGly 214

Qy       1253 CATTTCTATGCCAGTCAGATGGCTCAACGGAAGATCCGAGACATCCTGGCCCAGGTTAAG 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 HisPheTyrAlaSerGlnMetAlaGlnArgLysIleArgAspIleLeuAlaGlnValLys 234

Qy       1313 CAGCAGCATCAGAAGGGACAGAGTAACCAGGCCCAGGCACGGAGGAAG 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 GlnGlnHisGlnLysGlyGlnSerAsnGlnAlaGlnAlaArgArgLys 250

RESULT 7
AAB11328
ID   AAB11328 standard; protein; 579 AA.
XX
AC   AAB11328;
XX
DT   21-FEB-2001  (first entry)
XX
DE   Human lung cancer-associated protein L523S.
XX
KW   Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW   vaccine; detection.
XX
OS   Homo sapiens.
XX
PN   WO200061612-A2.
XX
PD   19-OCT-2000.
XX
PF   03-APR-2000; 2000WO-US008896.
XX
PR   02-APR-1999;   99US-00285479.
PR   17-DEC-1999;   99US-00466396.
PR   30-DEC-1999;   99US-00476496.
PR   10-JAN-2000; 2000US-00480884.
PR   22-FEB-2000; 2000US-00510376.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Fan L;
XX
DR   WPI; 2000-628399/60.
DR   N-PSDB; AAC65900.
XX
PT   Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT   protein is used for detecting and monitoring progression of lung cancer
PT   in a patient.
XX
PS   Claim 3; Page 186-188; 261pp; English.
XX
CC   This invention describes a novel isolated polypeptide (I) which
CC   comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC   which have cytostatic activity. The polypeptides and polynucleotides are
CC   used in compositions and vaccines to inhibit the development of cancer,
CC   especially lung cancer, in a patient. Methods described in the invention
CC   can be used to monitor the progression of a cancer by carrying out the
CC   detection at subsequent time points and comparing the results from the
CC   different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC   are treated with P2, polynucleotides encoding P2 or antigen presenting
CC   cells expressing P2 and then administered to the patient to inhibit
CC   development of cancer
XX
SQ   Sequence 579 AA;
```

```
Alignment Scores:
Pred. No.:              3.86e-26        Length:         579
Score:                  36.00           Matches:        36
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            6.43%           Indels:         0
DB:                     3               Gaps:           0

US-09-270-437D-5 (1-1708) x AAB11328 (1-579)

Qy        254 GTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAG 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226

Qy        314 TCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAA 361
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242

RESULT 8
AAB11365
ID   AAB11365 standard; protein; 579 AA.
XX
AC   AAB11365;
XX
DT   21-FEB-2001  (first entry)
XX
DE   Human lung cancer associated antigen L523S.
XX
KW   Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW   vaccine; detection.
XX
OS   Homo sapiens.
XX
PN   WO200061612-A2.
XX
PD   19-OCT-2000.
XX
PF   03-APR-2000; 2000WO-US008896.
XX
PR   02-APR-1999;   99US-00285479.
PR   17-DEC-1999;   99US-00466396.
PR   30-DEC-1999;   99US-00476496.
PR   10-JAN-2000; 2000US-00480884.
PR   22-FEB-2000; 2000US-00510376.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Fan L;
XX
DR   WPI; 2000-628399/60.
DR   N-PSDB; AAC66035.
XX
PT   Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT   protein is used for detecting and monitoring progression of lung cancer
PT   in a patient.
XX
PS   Claim 3; Page 259-261; 261pp; English.
XX
CC   This invention describes a novel isolated polypeptide (I) which
CC   comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC   which have cytostatic activity. The polypeptides and polynucleotides are
CC   used in compositions and vaccines to inhibit the development of cancer,
CC   especially lung cancer, in a patient. Methods described in the invention
CC   can be used to monitor the progression of a cancer by carrying out the
CC   detection at subsequent time points and comparing the results from the
CC   different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC   are treated with P2, polynucleotides encoding P2 or antigen presenting
CC   cells expressing P2 and then administered to the patient to inhibit
CC   development of cancer
XX
SQ   Sequence 579 AA;

Alignment Scores:
Pred. No.:              3.86e-26        Length:         579
Score:                  36.00           Matches:        36
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            6.43%           Indels:         0
DB:                     3               Gaps:           0

US-09-270-437D-5 (1-1708) x AAB11365 (1-579)

Qy        254 GTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAG 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226

Qy        314 TCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAA 361
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242

RESULT 9
ABB74960
ID   ABB74960 standard; protein; 579 AA.
XX
AC   ABB74960;
XX
DT   01-MAY-2002  (first entry)
XX
DE   Human lung tumour L523S protein sequence SEQ ID NO:176.
XX
KW   Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
KW   immune response.
XX
OS   Homo sapiens.
XX
PN   WO200200174-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US021065.
XX
PR   28-JUN-2000; 2000US-00606421.
PR   02-AUG-2000; 2000US-00630940.
PR   21-AUG-2000; 2000US-00643597.
PR   15-SEP-2000; 2000US-00662786.
PR   09-OCT-2000; 2000US-00685696.
PR   12-DEC-2000; 2000US-00735705.
PR   07-MAY-2001; 2001US-00850716.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Wang A,  Skeiky YAW,  Li SX,  Kalos MD,  Henderson RA;
PI   Mcneill PD,  Fanger N,  Retter MW,  Marnerakis M,  Fanger GR;
PI   Vedvick TS,  Carter D,  Watanabe Y,  Peckham DW;
XX
DR   WPI; 2002-090513/12.
DR   N-PSDB; ABL49119.
XX
PT   Polynucleotides encoding lung tumor polypeptides, useful for treating
PT   lung cancer or stimulating an immune response.
XX
PS   Example 2; Page 267-268; 374pp; English.
XX
CC   The present invention describes human lung tumour proteins. Human lung
CC   tumour proteins and polynucleotides have cytostatic and immunostimulant
CC   activities, and can be used in vaccine production. Compositions
CC   comprising the lung tumour proteins, polynucleotides, antibodies, fusion
CC   proteins, T cell populations, or antigen presenting cells that express
CC   the lung tumour proteins are useful for treating lung cancer or
CC   stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
CC   ABB75070 represent sequences used in the exemplification of the present
CC   invention
XX
SQ   Sequence 579 AA;

Alignment Scores:
```

```
Pred. No.:              3.86e-26        Length:         579
Score:                  36.00           Matches:        36
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            6.43%           Indels:         0
DB:                     5               Gaps:           0

US-09-270-437D-5 (1-1708) x ABB74960 (1-579)

Qy        254 GTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAG 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226

Qy        314 TCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAA 361
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242

RESULT 10
ABB75053
ID   ABB75053 standard; protein; 579 AA.
XX
AC   ABB75053;
XX
DT   01-MAY-2002  (first entry)
XX
DE   Human lung tumour L523S recombinant protein sequence SEQ ID NO:446.
XX
KW   Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
KW   immune response.
XX
OS   Homo sapiens.
XX
PN   WO200200174-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US021065.
XX
PR   28-JUN-2000; 2000US-00606421.
PR   02-AUG-2000; 2000US-00630940.
PR   21-AUG-2000; 2000US-00643597.
PR   15-SEP-2000; 2000US-00662786.
PR   09-OCT-2000; 2000US-00685696.
PR   12-DEC-2000; 2000US-00735705.
PR   07-MAY-2001; 2001US-00850716.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Wang A,  Skeiky YAW,  Li SX,  Kalos MD,  Henderson RA;
PI   Mcneill PD,  Fanger N,  Retter MW,  Marnerakis M,  Fanger GR;
PI   Vedvick TS,  Carter D,  Watanabe Y,  Peckham DW;
XX
DR   WPI; 2002-090513/12.
DR   N-PSDB; ABL49297.
XX
PT   Polynucleotides encoding lung tumor polypeptides, useful for treating
PT   lung cancer or stimulating an immune response.
XX
PS   Claim 2; Page 365-367; 374pp; English.
XX
CC   The present invention describes human lung tumour proteins. Human lung
CC   tumour proteins and polynucleotides have cytostatic and immunostimulant
CC   activities, and can be used in vaccine production. Compositions
CC   comprising the lung tumour proteins, polynucleotides, antibodies, fusion
CC   proteins, T cell populations, or antigen presenting cells that express
CC   the lung tumour proteins are useful for treating lung cancer or
CC   stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
CC   ABB75070 represent sequences used in the exemplification of the present
CC   invention
XX
SQ   Sequence 579 AA;

Alignment Scores:
```

```
Pred. No.:              3.86e-26        Length:         579
Score:                  36.00           Matches:        36
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            6.43%           Indels:         0
DB:                     5               Gaps:           0

US-09-270-437D-5 (1-1708) x ABB75053 (1-579)

Qy        254 GTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAG 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226

Qy        314 TCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAA 361
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242

RESULT 11
ABB74997
ID   ABB74997 standard; protein; 579 AA.
XX
AC   ABB74997;
XX
DT   01-MAY-2002  (first entry)
XX
DE   Human lung tumour L523S protein sequence SEQ ID NO:348.
XX
KW   Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
KW   immune response.
XX
OS   Homo sapiens.
XX
PN   WO200200174-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US021065.
XX
PR   28-JUN-2000; 2000US-00606421.
PR   02-AUG-2000; 2000US-00630940.
PR   21-AUG-2000; 2000US-00643597.
PR   15-SEP-2000; 2000US-00662786.
PR   09-OCT-2000; 2000US-00685696.
PR   12-DEC-2000; 2000US-00735705.
PR   07-MAY-2001; 2001US-00850716.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Wang A,  Skeiky YAW,  Li SX,  Kalos MD,  Henderson RA;
PI   Mcneill PD,  Fanger N,  Retter MW,  Marnerakis M,  Fanger GR;
PI   Vedvick TS,  Carter D,  Watanabe Y,  Peckham DW;
XX
DR   WPI; 2002-090513/12.
DR   N-PSDB; ABL49254.
XX
PT   Polynucleotides encoding lung tumor polypeptides, useful for treating
PT   lung cancer or stimulating an immune response.
XX
PS   Example 2; Page 330-332; 374pp; English.
XX
CC   The present invention describes human lung tumour proteins. Human lung
CC   tumour proteins and polynucleotides have cytostatic and immunostimulant
CC   activities, and can be used in vaccine production. Compositions
CC   comprising the lung tumour proteins, polynucleotides, antibodies, fusion
CC   proteins, T cell populations, or antigen presenting cells that express
CC   the lung tumour proteins are useful for treating lung cancer or
CC   stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
CC   ABB75070 represent sequences used in the exemplification of the present
CC   invention
XX
SQ   Sequence 579 AA;

Alignment Scores:
```

Pred. No.:	3.86e-26	Length:	579
Score:	36.00	Matches:	36
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	6.43%	Indels:	0
DB:	5	Gaps:	0

US-09-270-437D-5 (1-1708) x ABB74997 (1-579)

```
Qy        254 GTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAG 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226

Qy        314 TCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAA 361
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242
```

RESULT 12
ABB75054
ID ABB75054 standard; protein; 579 AA.
XX
AC ABB75054;
XX
DT 01-MAY-2002 (first entry)
XX
DE Human lung tumour L523S recombinant protein sequence SEQ ID NO:449.
XX
KW Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
KW immune response.
XX
OS Homo sapiens.
XX
PN WO200200174-A2.
XX
PD 03-JAN-2002.
XX
PF 28-JUN-2001; 2001WO-US021065.
XX
PR 28-JUN-2000; 2000US-00606421.
PR 02-AUG-2000; 2000US-00630940.
PR 21-AUG-2000; 2000US-00643597.
PR 15-SEP-2000; 2000US-00662786.
PR 09-OCT-2000; 2000US-00685696.
PR 12-DEC-2000; 2000US-00735705.
PR 07-MAY-2001; 2001US-00850716.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
PI Mcneill PD, Fanger N, Retter MW, Marnerakis M, Fanger GR;
PI Vedvick TS, Carter D, Watanabe Y, Peckham DW;
XX
DR WPI; 2002-090513/12.
DR N-PSDB; ABL49299.
XX
PT Polynucleotides encoding lung tumor polypeptides, useful for treating
PT lung cancer or stimulating an immune response.
XX
PS Claim 2; Page 368-369; 374pp; English.
XX
CC The present invention describes human lung tumour proteins. Human lung
CC tumour proteins and polynucleotides have cytostatic and immunostimulant
CC activities, and can be used in vaccine production. Compositions
CC comprising the lung tumour proteins, polynucleotides, antibodies, fusion
CC proteins, T cell populations, or antigen presenting cells that express
CC the lung tumour proteins are useful for treating lung cancer or
CC stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
CC ABB75070 represent sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 579 AA;

Alignment Scores:

Pred. No.:	3.86e-26	Length:	579
Score:	36.00	Matches:	36
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	6.43%	Indels:	0
DB:	5	Gaps:	0

US-09-270-437D-5 (1-1708) x ABB75054 (1-579)

```
Qy        254 GTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAG 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226

Qy        314 TCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAA 361
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242
```

RESULT 13
ABP61917
ID ABP61917 standard; protein; 579 AA.
XX
AC ABP61917;
XX
DT 07-OCT-2002 (first entry)
XX
DE Human lung cancer associated protein sequence SEQ ID NO:348.
XX
KW Human; lung cancer; lung tumour; cytostatic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO200247534-A2.
XX
PD 20-JUN-2002.
XX
PF 30-NOV-2001; 2001WO-US047576.
XX
PR 12-DEC-2000; 2000US-00735705.
PR 07-MAY-2001; 2001US-00850716.
PR 28-JUN-2001; 2001US-00897778.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
PI Mcneill PD, Fanger N, Retter MW, Durham M, Fanger GR, Vedvick TS;
PI Carter D, Watanabe Y, Peckham DW, Cai F, Foy TM;
XX
DR WPI; 2002-583465/62.
DR N-PSDB; ABQ92440.
XX
PT Novel lung carcinoma polynucleotide sequences and polypeptides encoded by
PT the polynucleotides, useful in pharmaceutical compositions such as
PT vaccines and as markers to indicate the presence of lung cancer.
XX
PS Example 2; Page 337-339; 381pp; English.
XX
CC The present invention describes isolated human lung carcinoma
CC polynucleotides (I) and polypeptides (II). (I) and (II) have cytostatic
CC activity, and can be used in gene therapy and in vaccines. Compositions
CC comprising (I) or (II) can be used for stimulating an immune response in
CC a patient and for treating lung cancer in a patient. Oligonucleotides of
CC (I) can be used for detecting the presence of a cancer in a patient, by
CC obtaining a biological sample from the patient, contacting the biological
CC sample with the oligonucleotide, detecting in the sample, an amount of
CC polynucleotide that hybridises to the oligonucleotide and comparing the
CC amount of polynucleotide that hybridises to the oligonucleotide to a
CC predetermined cut-off value, and determining the presence of a cancer in
CC the patient. (I) and (II) are useful in pharmaceutical compositions, e.g.
CC vaccines. (I) is useful as a marker to indicate the presence or absence
CC of a cancer such as lung cancer. ABQ92145 to ABQ92486 and ABP61866 to
CC ABP61992 represent sequences used in the exemplification of the present
CC invention
XX

```
SQ   Sequence 579 AA;

Alignment Scores:
Pred. No.:              3.86e-26        Length:         579
Score:                  36.00           Matches:        36
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            6.43%           Indels:         0
DB:                     5               Gaps:           0

US-09-270-437D-5 (1-1708) x ABP61917 (1-579)

Qy        254 GTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAG 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226

Qy        314 TCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAA 361
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242

RESULT 14
ABP61974
ID   ABP61974 standard; protein; 579 AA.
XX
AC   ABP61974;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Human lung cancer associated protein sequence SEQ ID NO:449.
XX
KW   Human; lung cancer; lung tumour; cytostatic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200247534-A2.
XX
PD   20-JUN-2002.
XX
PF   30-NOV-2001; 2001WO-US047576.
XX
PR   12-DEC-2000; 2000US-00735705.
PR   07-MAY-2001; 2001US-00850716.
PR   28-JUN-2001; 2001US-00897778.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Wang A,  Skeiky YAW,  Li SX,  Kalos MD,  Henderson RA;
PI   Mcneill PD,  Fanger N,  Retter MW,  Durham M,  Fanger GR,  Vedvick TS;
PI   Carter D,  Watanabe Y,  Peckham DW,  Cai F,  Foy TM;
XX
DR   WPI; 2002-583465/62.
DR   N-PSDB; ABQ92485.
XX
PT   Novel lung carcinoma polynucleotide sequences and polypeptides encoded by
PT   the polynucleotides, useful in pharmaceutical compositions such as
PT   vaccines and as markers to indicate the presence of lung cancer.
XX
PS   Claim 9; Page 375-377; 381pp; English.
XX
CC   The present invention describes isolated human lung carcinoma
CC   polynucleotides (I) and polypeptides (II). (I) and (II) have cytostatic
CC   activity, and can be used in gene therapy and in vaccines. Compositions
CC   comprising (I) or (II) can be used for stimulating an immune response in
CC   a patient and for treating lung cancer in a patient. Oligonucleotides of
CC   (I) can be used for detecting the presence of a cancer in a patient, by
CC   obtaining a biological sample from the patient, contacting the biological
CC   sample with the oligonucleotide, detecting in the sample, an amount of
CC   polynucleotide that hybridises to the oligonucleotide and comparing the
CC   amount of polynucleotide that hybridises to the oligonucleotide to a
CC   predetermined cut-off value, and determining the presence of a cancer in
CC   the patient. (I) and (II) are useful in pharmaceutical compositions, e.g.
CC   vaccines. (I) is useful as a marker to indicate the presence or absence
CC   of a cancer such as lung cancer. ABQ92145 to ABQ92486 and ABP61866 to
CC   ABP61992 represent sequences used in the exemplification of the present
CC   invention
XX
SQ   Sequence 579 AA;

Alignment Scores:
Pred. No.:              3.86e-26        Length:         579
Score:                  36.00           Matches:        36
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            6.43%           Indels:         0
DB:                     5               Gaps:           0

US-09-270-437D-5 (1-1708) x ABP61974 (1-579)

Qy        254 GTGGGTGCCATTATTGGCAAGGAGGGGGCCACCATCCGCAACATCACAAAACAGACCCAG 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ValGlyAlaIleIleGlyLysGluGlyAlaThrIleArgAsnIleThrLysGlnThrGln 226

Qy        314 TCCAAGATAGACGTGCATAGGAAGGAGAACGCAGGTGCAGCTGAAAAA 361
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        227 SerLysIleAspValHisArgLysGluAsnAlaGlyAlaAlaGluLys 242

RESULT 15
ABP61880
ID   ABP61880 standard; protein; 579 AA.
XX
AC   ABP61880;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Human lung cancer associated protein sequence SEQ ID NO:176.
XX
KW   Human; lung cancer; lung tumour; cytostatic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200247534-A2.
XX
PD   20-JUN-2002.
XX
PF   30-NOV-2001; 2001WO-US047576.
XX
PR   12-DEC-2000; 2000US-00735705.
PR   07-MAY-2001; 2001US-00850716.
PR   28-JUN-2001; 2001US-00897778.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Wang A,  Skeiky YAW,  Li SX,  Kalos MD,  Henderson RA;
PI   Mcneill PD,  Fanger N,  Retter MW,  Durham M,  Fanger GR,  Vedvick TS;
PI   Carter D,  Watanabe Y,  Peckham DW,  Cai F,  Foy TM;
XX
DR   WPI; 2002-583465/62.
DR   N-PSDB; ABQ92305.
XX
PT   Novel lung carcinoma polynucleotide sequences and polypeptides encoded by
PT   the polynucleotides, useful in pharmaceutical compositions such as
PT   vaccines and as markers to indicate the presence of lung cancer.
XX
PS   Example 2; Page 274-275; 381pp; English.
XX
CC   The present invention describes isolated human lung carcinoma
CC   polynucleotides (I) and polypeptides (II). (I) and (II) have cytostatic
CC   activity, and can be used in gene therapy and in vaccines. Compositions
CC   comprising (I) or (II) can be used for stimulating an immune response in
CC   a patient and for treating lung cancer in a patient. Oligonucleotides of
CC   (I) can be used for detecting the presence of a cancer in a patient, by
CC   obtaining a biological sample from the patient, contacting the biological
CC   sample with the oligonucleotide, detecting in the sample, an amount of
CC   polynucleotide that hybridises to the oligonucleotide and comparing the
CC   amount of polynucleotide that hybridises to the oligonucleotide to a
CC   predetermined cut-off value, and determining the presence of a cancer in
```